

```
/ STRAIN: NRRL 2338
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 744..6659
/ OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
/ OTHER INFORMATION: MODULE 1"
/ OTHER INFORMATION: /label= FUNCTION
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 744..11219
/ OTHER INFORMATION: /function= "gene= "eryA""
/ OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
/ OTHER INFORMATION: 6-deoxyerythronolide B""
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 744..1868
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyltransferase domain 1 of module 1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1998..2198
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyl carrier domain 1 of module 1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 2250..3626
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 3831..4811
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyltransferase 2 domain of module 1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 5574..6125
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoreductase domain of module 1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 6369..6626
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyl carrier domain 2 of module 1"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 6678..11219
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: module 2"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 6678..8066
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 8262..9305
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyltransferase domain of module 2"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 9906..10454
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: beta-ketoreductase domain of module 2"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 10707..10964
/ OTHER INFORMATION: /function= "approximate span of
/ OTHER INFORMATION: acyl carrier domain of module 2"
/ ;
```

US-08-439-009A-1

```
Query Match          38.5%; Score 94.6; DB 3; Length 11219;
Best Local Similarity 61.6%; Pred. No. 3.4e-17;
Matches 151; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY      1  CTGGCCCATCAGACCTCGGCCGACCAACGGGCGCGCATTCGCTCGAGCTCGTACGAGACCAT 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10707  CTGGCGGGTCTGCCCGGTGCCGAGCGGACGGCGGAGCTGGTGGCTTGTCCGCCACGACG 10766
QY      61  GTGGCGGAGTGTCTCCGGCACCGCGACCGGAAAGCCATCGGCCCGGACGACGACGTCGTTCCGT 120
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10767  ACCGCGACCGTGTCTGGGCCACGACGACCGGAGGGGGTCCGGCGACCAACGCGCTTCAAG 10826
QY      121  GCACTCGGCTTCGATTCACTACGGCGCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10827  GAGCTCGGGTTGACTCGCTGGCGCGCGTCCGGCTGCGCAACTGCTCAACGGCGGCACC 10886
QY      181  GGACTCCGCTTCTCTGCTGCTGCTTTCGACCAACCGGACCCCTGCCCCAAACTCGCCGTA 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10887  GGGCTCCGCTGCTGCGCGCTGCTTCGAGACCACCGGACGCGCTCGCGCGTTCGCGGT 10946
QY      241  CACCT 245
      |||
Db      10947  TTCTT 10951
```

Search completed: June 18, 2003, 01:07:08
Job time : 20.6565 secs

```
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Saccharopolyspora erythraea
;; STRAIN: NRRL 2338
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 744..6659
;; OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
;; OTHER INFORMATION: MODULE 1"
;; OTHER INFORMATION: /label= FUNCTION
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 744..11219
;; OTHER INFORMATION: /function= "gene= "eryA"
;; OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
;; OTHER INFORMATION: 6-deoxyerythronolide B"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 744..1868
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: acyltransferase domain 1 of module 1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 1998..2198
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: acyl carrier domain 1 of module 1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 2250..3626
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 3831..4811
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: acyltransferase 2 domain of module 1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 5574..6125
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: beta-ketoreductase domain of module 1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 6369..6626
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: acyl carrier domain 2 of module 1"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 6678..11219
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: module 2"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 6678..8066
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 8262..9305
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: acyltransferase domain of module 2"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 9906..10454
;; OTHER INFORMATION: /function= "approximate span of
;; OTHER INFORMATION: beta-ketoreductase domain of module 2"
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: 10707..10964
;; OTHER INFORMATION: /function= "approximate span of
```

```
;; OTHER INFORMATION: acyl carrier domain of module 2"
;; US-07-642-734C-1
;; Query Match 38.5%; Score 94.6; DB 1; Length 11219;
;; Best Local Similarity 61.8%; Pred. No. 3.4e-17;
;; Matches 151; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGGCCGACCAACGGGCGCGATTGCTCGAGCTCGTACGAGACCAT 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10707 CTGGCGGGTCTGCGCGTGCAGCGGAGCGGAGCTGGTGGCTGTGCTCCGACGAGC 10766
QY 61 GTGGCGGAGTGTCTCCGCAACGCGGACCCGAAAGCATCGCGCCGACCAAGTGTTCGCT 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10767 ACCGCGACCGTGTGGGCCAGACGACCCGAAAGCGGTGCGCGGACCGCGCGTTCAAG 10826
QY 121 GCACCTCGGCTTCGATTCACTCAGCGCGGTGAGTTCGAAACCTGTGATCAAGGCAACA 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10827 GAGCTCGGTTGCACTCGTGGCGCGCTCGGCTCGCAACCTGTGATCAAGCGGCAAC 10886
QY 181 GGACTCGGCTTCCTGCTCGTGTGCTTCGACACCGCGACCCCTGCCAAACTCGCGGTA 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10887 GGGCTCGGCTGCGGTGCGAGGCTGGTCTTCGACCAACCGGCTCCGCGGTGCGCGGT 10946
QY 241 CACCT 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10947 TTCCT 10951

RESULT 15
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
```



```

; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R.
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

Query Match 40.8%; Score 100.4; DB 3; Length 80161;
Best Local Similarity 63.0%; Pred. No. 1.3e-18;
Matches 155; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
DB 39790 CTCGGGAGTGC CGGAGCCCAACGACGACCACTGGTGTGTCGGCCAG 39849

QY 61 GTGGGGCAGTGTCTCGGACCGCGGACCGGAAAGCCATCGCGCCGACGAGTGTTCCT 120
DB 39850 GTGGCAACCGTCTCGGGCACCGGAGTCCGAGGAGTCCAGCCGAGGGGGTTCGCG 39909

QY 121 GCACCTGGTTCGATTCACTACGCGCGTGTGTCGAAACCTGCTGATCAAGGCAACA 180
DB 39910 GCGCTGGGTTCGACTCCCTCATGCGGTGGATCTGCGCAATCGTTTGACCAACGCGCAC 39969

QY 181 GGACTCGGCTTCCTGCTCGCTGTCTCGACCAACCGCTTTCGACTACCGGCGGTA 240
DB 39970 GGGTTGCGGCTCGCGACCAACCGCTTTCGACTACCGGCGGCGCTTGGCGGCT 40029

QY 241 CACCTG 246
DB 40030 CACCTG 40035

RESULT 11
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-34

Query Match 40.0%; Score 98.4; DB 4; Length 4689;
Best Local Similarity 64.5%; Pred. No. 2.6e-18;
Matches 147; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

```

; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

Query Match 40.8%; Score 100.4; DB 4; Length 80161;
Best Local Similarity 63.0%; Pred. No. 1.3e-18;
Matches 155; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
DB 39790 CTCGGGAGTGC CGGAGCCCAACGACGACCACTGGTGTGTCGGCCAG 39849

QY 61 GTGGGGCAGTGTCTCGGACCGCGGACCGGAAAGCCATCGCGCCGACGAGTGTTCCT 120
DB 39850 GTGGCAACCGTCTCGGGCACCGGAGTCCGAGGAGTCCAGCCGAGGGGGTTCGCG 39909

QY 121 GCACCTGGTTCGATTCACTACGCGCGTGTGTCGAAACCTGCTGATCAAGGCAACA 180
DB 39910 GCGCTGGGTTCGACTCCCTCATGCGGTGGATCTGCGCAATCGTTTGACCAACGCGCAC 39969

QY 181 GGACTCGGCTTCCTGCTCGCTGTCTCGACCAACCGCTTTCGACTACCGGCGGTA 240
DB 39970 GGGTTGCGGCTCGCGACCAACCGCTTTCGACTACCGGCGGCGCTTGGCGGCT 40029

QY 241 CACCTG 246
DB 40030 CACCTG 40035

```

CITY: INDIANAPOLIS
 STATE: IN
 COUNTRY: USA
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCII(DOS) Text only
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804,227C
 FILING DATE: February 21, 1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Plant, Thomas, G.
 REGISTRATION NUMBER: 35,784
 REFERENCE/DOCKET NUMBER: X-8231
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-2459
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 44377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 350..14002
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 14046..20036
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20110..31284
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 31329..36071
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36155..41830
 US-08-804-227C-7

Query Match 40.9%; Score 100.6; DB 2; Length 44377;
 Best Local Similarity 65.2%; Pred. No. 1e-18;
 Matches 148; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 20 CCAGCAACGGCGCGCATTCCTCGAGCTCTGACGACCATGTGGCGGAGTCTCGGC 79
 Db 19495 CCGAGCGTACGGCTCTCTCTCGACCTGTGGGACCGAGTCTGGCGGCTCGGAC 19554
 QY 80 ACAGCGACCGGAAGCCATCGCCCGACCGAGTCTCGTGCACCTCGGCTTCGATTCCAC 139
 Db 19555 ACAGCGACCGCGCGGATCGCGCGCCCGGACCTTCAAGGACGGCGGATTCGATCTCC 19614
 QY 140 TCACGGCGTGCAGTTCGGAACCTGCTGATCAAGGCAACAGGACTCCGCTTCCTGTCT 199
 Db 19615 TCACCGTGTGACCTCCGCAACCGGCTGAACACACGACCGGACTCGGCTCCCGCA 19674
 QY 200 CGCTGGTCTTCGACCGACCGACCCCTGCGCAAACTCGCGGTACACCTG 246
 Db 19675 CCCTGCTTTCGACCGACCGGCTCGCCCTCGCGCAACTCTG 19721

RESULT 8

US-08-804-198-1
 ; Sequence 1, Application US/08804198
 ; Patent No. 5945320
 ; GENERAL INFORMATION:
 ; APPLICANT: Burgett, Stanley G.
 ; APPLICANT: Kuhstoss, Stuart A.
 ; APPLICANT: Rao, Nagaraja R.
 ; APPLICANT: Richardson, Mark A.

APPLICANT: Rosteck, Paul R., Jr.
 TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PAUL R. CANTRELL 1138
 STREET: LILLY CORPORATE CENTER
 CITY: INDIANAPOLIS
 STATE: IN
 COUNTRY: USA
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Macintosh
 OPERATING SYSTEM: Macintosh 7.0
 SOFTWARE: Microsoft Word 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804,198
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CANTRELL, PAUL R.
 REGISTRATION NUMBER: 36,470
 REFERENCE/DOCKET NUMBER: P9113
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-3885
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 44377 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 350..14002
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 14046..20036
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20110..31284
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 31329..36071
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36155..41830
 US-08-804-198-1

Query Match 40.9%; Score 100.6; DB 2; Length 44377;
 Best Local Similarity 65.2%; Pred. No. 1e-18;
 Matches 148; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 20 CCAGCAACGGCGCGCATTCCTCGAGCTCTGACGACCATGTGGCGGAGTCTCGGC 79
 Db 19495 CCGAGCGTACGGCTCTCTCTCGACCTGTGGGACCGAGTCTGGCGGCTCGGAC 19554
 QY 80 ACAGCGACCGGAAGCCATCGCCCGACCGAGTCTCGTGCACCTCGGCTTCGATTCCAC 139
 Db 19555 ACAGCGACCGCGCGGATCGCGCGCCCGGACCTTCAAGGACGGCGGATTCGATCTCC 19614
 QY 140 TCACGGCGTGCAGTTCGGAACCTGCTGATCAAGGCAACAGGACTCCGCTTCCTGTCT 199
 Db 19615 TCACCGTGTGACCTCCGCAACCGGCTGAACACACGACCGGACTCGGCTCCCGCA 19674
 QY 200 CGCTGGTCTTCGACCGACCGACCCCTGCGCAAACTCGCGGTACACCTG 246
 Db 19675 CCCTGCTTTCGACCGACCGGCTCGCCCTCGCGCAACTCTG 19721

RESULT 9

US-09-036-987A-1
 ; Sequence 1, Application US/09036987A

```

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/804,227C
  FILING DATE: February 21, 1997
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Plant, Thomas, G.
  REGISTRATION NUMBER: 35,784
  REFERENCE/DOCKET NUMBER: X-8231
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 317-276-2459
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 43280 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 816..14234
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 14351..19945
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 20010..31199
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 31232..36067
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 36249..41774
  US-08-804-227C-1

Query Match 41.5%; Score 102; DB 2; Length 43280;
Best Local Similarity 63.4%; Pred. No. 4.2e-19;
Matches 156; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 1 CTGGCCCATCAGACGCTCGGCCGACCAACGGGCGCATTCCTCGAGCTCGTAGGACCAT 60
Db 30648 CTACCGGGGCTCACGGCACAGGAAGGGCAGCGCCCTCGTGGCCGAGATCGCGCCAC 30707
Qy 61 GTGGCGGAGTGTCTCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCGAGTCGTTCCGT 120
Db 30708 GCGGCGGCGGTGCTGGGCGACCGCTCCGACGACTGATCCCCGAGACCGGGCTTCAAG 30767
Qy 121 GCACCTCGGTTTCGATTCATCTACGCGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCACA 180
Db 30768 GACCTCGGTTTCGATTCGCTACCGCGCTGGAGATGCGCAACCGGCTGAGCGCGCCACC 30827
Qy 181 GGACTCGGCTTCTGTCTCGCTGTGCTTCGACACCGGACCCCTGCCAAACTCGCCGTA 240
Db 30828 GGCCTCGGCTCGCGCCACCCCTCGTCTTCGACACCGGACCCCGGCGAGCTGGCCGCG 30887
Qy 241 CACCTG 246
Db 30888 CACCTG 30893

RESULT 7
US-08-804-227C-7
  Sequence 7, Application US/08804227C
  Patent No. 5876991
  GENERAL INFORMATION:
    APPLICANT: DeHoff, Bradley S.
    APPLICANT: Kuhstoss, Stuart A.
    APPLICANT: Rostock, Paul R., Jr.
    APPLICANT: Sutton, Kimberly L.
    TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
    NUMBER OF SEQUENCES: 15
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: THOMAS G. PLANT 1501
      STREET: LILLY CORPORATE CENTER

```


Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	110.2	44.8	50937	4	US-09-428-517-1	Sequence 1, Appli
2	103.6	42.1	1581	4	US-09-434-288-7	Sequence 7, Appli
3	103.6	42.1	11220	4	US-09-105-537-32	Sequence 32, Appli
4	103.6	42.1	36778	4	US-09-105-537-5	Sequence 5, Appli
5	103.6	42.1	38506	3	US-09-320-878-19	Sequence 19, Appli
6	102	41.5	43280	2	US-08-804-227C-1	Sequence 1, Appli
7	100.6	40.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
8	100.6	40.9	44377	2	US-08-804-198-1	Sequence 1, Appli
9	100.4	40.8	80161	3	US-09-036-987A-1	Sequence 1, Appli
10	100.4	40.8	80161	3	US-09-370-700-1	Sequence 1, Appli
11	98.4	39.0	4689	4	US-09-105-537-34	Sequence 34, Appli
12	97.8	39.8	1434	4	US-09-434-288-3	Sequence 3, Appli
13	96.2	39.1	15872	4	US-09-105-537-1	Sequence 1, Appli
14	94.6	38.5	11219	1	US-07-642-734C-1	Sequence 1, Appli
15	94.6	38.5	11219	3	US-08-439-009A-1	Sequence 1, Appli
16	94.6	38.5	13842	4	US-09-105-537-30	Sequence 30, Appli
17	92.8	37.7	561	3	US-09-154-083-13	Sequence 13, Appli
18	91	37.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
19	91	37.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
20	89	36.2	13987	2	US-08-804-227C-13	Sequence 13, Appli
21	86	35.0	4403765	4	US-09-103-840A-1	Sequence 2, Appli
22	86	35.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
23	84.4	34.3	28958	1	US-08-258-261B-6	Sequence 6, Appli
24	84.4	34.3	28958	1	US-08-456-837-6	Sequence 6, Appli
25	84.4	34.3	28958	1	US-08-457-342-6	Sequence 6, Appli
26	84.4	34.3	28958	1	US-08-457-464A-6	Sequence 6, Appli
27	84.4	34.3	28958	1	US-08-458-076A-6	Sequence 6, Appli

THIS PAGE BLANK (uspro)

PR 06-MAY-1998; 98US-0073538.
XX 28-AUG-1998; 98US-0141908.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX WPI; 2000-610844/58.
XX
XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT
XX
XX Disclosure; Columns 15-32; 117pp; English.
XX
XX The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 42.1%; Score 103.6; DB 21; Length 38506;
Best Local Similarity 63.8%; Pred. No. 2.5e-18;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTGCGCGGACCAACGGCGCGCATTTGCTGAGCTCGTAGGACCAT 60
DB 24507 CTGCGCGCATGACACCGGACGACCGGGTGGCGCACCTGGGGACCTCGTCGTACGCAC 24566

QY 61 GTGGCGGCGAGTGTCTCGGACGCGGACCCGGAAGCCATCGCGCCCGACGATCGTTCCGT 120
DB 24567 GTGGCGACCGTCTGGGACACGCGCACCCGCGGGTGGACCTGGAGCGGGCTTCCGC 24626

QY 121 GCACTCGGCTTCGATTCTACTACGGCGGTCGAGTTCGGAACCTGCTGATCAAGGCAACA 180
DB 24627 GACACCGGTTTGACTCGCTACCGCGGTCGAACTCCGCAACCGTCTCAACGCGCGACC 24686

QY 181 GGACTCCGCTTCCTGCTCTCGTGTCTTCGACCAACCCGACCCCTGCCAAACTCGCCGTA 240
DB 24687 GGGCTGCGGTGCGGSCCAGCTGGTCTTCGACCACCCACCCCGGGGAGCTCGCCGGG 24746

QY 241 CACCTG 246
DB 24747 CACCTG 24752

Search completed: June 17, 2003, 13:48:30
Job time : 51.0269 secs

XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin -
XX Claim 15; Page 398-403; 438pp; English.
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AA287295-287302 represent macrolide biosynthetic genes from
CC Streptomyces venezuelae ATCC 15439, which encode proteins
CC AA287190-287197.
XX SQ Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;
Query Match 42.1%; Score 103.6; DB 21; Length 11220;
Best Local Similarity 63.8%; Pred. No. 2.1e-18;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGCGCGACCAACGGGCGGCATTTGCTCGAGTGTGTTACGAGACCAT 60
Db 10678 CTGCGCGCATGACACCGGACGACCGCGGTGCGGCACCTGCGGACCTGCTCCGTACGCCAC 10737
QY 61 GTGCGGCGAGTGTCTCGGCGACGGGACCGCCGAAGCCATCGCGCCGACGAGTGTTCGGT 120
Db 10738 GTGCGGCGAGTGTCTCGGCGACGGGACCGCCGAGCGCGGTGACCTGAGCGGGGCTTCCGC 10797
QY 121 GCACTCGGCTTCGATTTCACCTCAGCGCGCGTCTGAGTTCGGAACCTGCTGATCAAGGCAACA 180
Db 10798 GACACCGGTTTCGACTGCTCACCAGCGGTGCACTCGGACCGTCTCAAGCGCGGAC 10857
QY 181 GGACTCGGCTTCCTGTCTGCTGTGTTGAGTTCGAGACCCCGACCCCTCCCAACTCGCCGTA 240
Db 10858 GGGCTCGGCTGCGCGCACGCTGCTTTCGACCAACCCCGCGGGAGCTCGCGCGG 10917
QY 241 CACCTG 246
Db 10918 CACCTG 10923
RESULT 13
AA287318
ID AA287318 standard; DNA; 36778 BP.
XX
AC AA287318;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;

KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX Streptomyces venezuelae ATCC15439.
XX Key Location/Qualifiers
FH 1742..15583
FT CDS
FT /tag= a
FT /product= "Pik gene cluster protein #1 (AA2877200)"
FT CDS
FT 15688..26907
FT /tag= b
FT /product= "Pik gene cluster protein #2 (AA2877201)"
FT CDS
FT 26991..31679
FT /tag= c
FT /product= "Pik gene cluster protein #3 (AA2877202)"
FT CDS
FT 31782..35822
FT /tag= d
FT /product= "Pik gene cluster protein #4 (AA2877203)"
FT CDS
FT 35819..36664
FT /tag= e
FT /product= "Pik gene cluster protein #4 (AA28780997)"
XX WO200000620-A2.
XX 06-JAN-2000.
XX 25-JUN-1999; 99WO-US14398.
XX 26-JUN-1998; 98US-0105537.
XX (MINU) UNIV MINNESOTA.
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX WPI; 2000-160679/14.
XX P-PSDB; AA2877200, AA2877201, AA2877202, AA2877203, AA28780997.
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT synthesis of methymycin and pikromycin -
XX Disclosure; Figure 31; 438pp; English.
XX The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 31.
XX Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

XX The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
 XX
 SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
 Query Match 42.8%; Score 105.2; DB 22; Length 65140;
 Best Local Similarity 64.2%; Pred. No. 1e-18;
 Matches 158; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCCGATTCGTCGAGCTCGTACGAGACCAT 60
 DB CTGCGCGGCTCGACCGCGCGCGCGCGGAGCCCTCTGGAATCGTCCGACCCAG 21556
 QY 61 GTGGGGGAGTGTCCTCGGACGCGGACCGGACCGGAAAGCCATCGCGCCGACGATGTTCCGT 120
 DB ATGCGCCAGGTCCTCGGACGCGGACCGGACCGGAGGTGAGACCGCGCGCCAGTTCAG 21616
 QY 121 GCACCTGGCTTCGATTCACCTCACGGCGCTCGAGTTCGGAACCTGTCGATCAAGGCAACA 180
 DB GACCTGGCTTCGATTCACCTCACGGCGCTCGAGTTCGGAACCTGTCGATCAAGGCAACA 21676
 QY 181 GGAACCTGGCTTCGATTCACCTCACGGCGCTCGAGTTCGGAACCTGTCGATCAAGGCAACA 240
 DB GGCTGGGCTGGCGCGCCGACCATGTTGTCGACTACCGACACACCGCCCTCGCGGAC 21736
 QY 241 CACCTG 246
 DB 21737 CACCTG 21742

RESULT 11

AA518438
 ID AA518438 standard; DNA; 1681 BP.
 AC AA518438;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Contig 115 DNA encoding S. narbonensis polyketide synthase.
 KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
 KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethylmycin;
 KW agriculture; ds.
 XX
 OS Streptomyces narbonensis.
 XX US6303767-B1.
 XX
 PD 16-OCT-2001.
 XX
 PF 05-NOV-1999; 99US-0434288.
 XX
 PR 05-NOV-1998; 98US-107093P.
 PR 27-MAY-1999; 99US-0320878.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.

XX Betlach MC, McDaniel R;
 XX WPI; 2002-065495/09.
 DR
 XX Nucleic acids encoding narbonolide polyketide synthases from
 PT Streptomyces narbonensis, useful for the recombinant production of
 PT polyketides, e.g. narbomycin -
 XX
 XX Claim 1; Column 18; 24pp; English.
 PS
 XX The present invention relates to recombinant DNA vectors (cosmids)
 CC that encode for the narbonolide polyketide synthase (PKS) enzyme and

CC various narbomycin modification enzymes from Streptomyces narbonensis.
 CC The recombinant DNA vectors can be used to produce recombinant ketide
 CC synthases and a variety of different polyketides (e.g. erythromycin,
 CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
 CC neomethylmycin) for use in agriculture, medicine and health. The
 CC recombinant vectors may be used to produce polyketides in relatively
 CC high yields. AA518432-AA518443 represent contig DNA sequences that
 CC encode for S. narbonensis PKS enzymes.

XX SQ Sequence 1681 BP; 231 A; 672 C; 580 G; 198 T; 0 other;
 Query Match 42.1%; Score 103.6; DB 24; Length 1681;
 Best Local Similarity 63.8%; Pred. No. 1.5e-18;
 Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTCGTCGAGCTCGTACGAGACCAT 60
 DB CTGCGCGGATACCCCGGACGACAGGACCGCGGACCTCGCGGACCTCGTACGAGCAC 600
 QY 61 GTGGGGGAGTGTCCTCGGACGCGGACCGGAAAGCCATCGCGCCGACGATGTTCCGT 120
 DB GTGGGGGAGTGTCCTCGGACGCGGACCGGCGCCCGGAGCTCGAGCGCGCTTCGCG 660
 QY 121 GCACCTGGCTTCGATTCACCTCACGGCGCTCGAGTTCGGAACCTGTCGATCAAGGCAACA 180
 DB GACACCGGTTTCGACTCCCTCACCGCGCTCGAGTTCGGAACCTGTCGATCAAGGCGCAC 720
 QY 181 GCACCTGGCTTCGATTCACCTCACGGCGCTCGAGTTCGGAACCTGTCGATCAAGGCGCAC 240
 DB GGCTGGGCTGGCGCGCCGACCATGTTGTCGACTACCGACACACCGCCCTCGCGGAC 780
 QY 241 CACCTG 246
 DB 781 CACCTG 786

RESULT 12

AAZ87298
 ID AAZ87298 standard; DNA; 11220 BP.
 XX
 AC AAZ87298;
 XX
 DT 05-JUN-2000 (first entry)
 XX
 DE S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
 KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethylmycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolaemia; crop protection agent; ds.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX
 FH Key Location/Qualifiers
 CDS 1..111220
 FT /*tag= a
 FT /product= "PikAII"
 XX
 PN WO200000620-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US14398.
 XX
 PR 26-JUN-1998; 98US-0105537.
 XX
 PA (MINU) UNIV MINNESOTA.
 XX
 PI Sherman DH, Liu H, Xue Y, Zhao L;
 XX WPI; 2000-160679/14.
 DR P-PSDB; AAY77193.

```

PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOWA ) KYOWA HAKKO KOGYO KK.
PA (KITA ) KITASATO INST.
XX
PI Endo H, Yanaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI; 2001-582053/65.
DR P-PSDB; AAG65266, AAG65267.
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
XX Disclosure; Page 103-149; 257pp; Japanese.
XX
XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.
XX
SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
Query Match 43.7%; Score 107.4; DB 22; Length 31422;
Best Local Similarity 64.9%; Pred. No. 2.3e-19;
Matches 159; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGCGGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60.
DB 20089 CTGGCGGCGAGACACAGACAGACACACACCATCTCTCGCTCTGCTCCCTCCAC 20148
QY 61 GTGGCGGCGAGTGTCTCGGCGACGGGACCCGAAAGCCATCGCGCGGACGAGTGTTCGGT 120
DB 20149 ATCGCCACCGTCTCGGCGACACACCGCGGACACCATCCCCCGACCGCGGTTCGGC 20208
QY 121 GCACCTGGCTTCGATTCACACGCGGCGTGTGAGTTCGGAACCTGCTGATCAAGGCAACA 180
DB 20209 GACCTCGGCTTCGACTCTCTCACCGCGGTGGAACCTGCAACCGCGCTCTCCGACCA 20268
QY 181 GGACTCGGCTTCCTGTCTGCTGCTGCTTCGACACCGCGGACCCCTCGCAACCTCGCGCGTA 240
DB 20269 GGACTCGGCTTCCTCGGCGGACCCCTCGGCTTCGACACCGCGGACCCCTCGCAAC 20328
QY 241 CACCT 245
DB 20329 CACCT 20333

RESULT 10
AADI17184
ID AADI17184 standard; DNA; 65140 BP.
XX
AC AADI17184;
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
XX Polypeptide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nys1; ds.
XX
XX Streptomyces noursei.
XX
XX Key Location/Qualifiers
XX CDS complement (1..1035)
FT /tags a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT

```

```

FT CDS complement (1056..2576)
FT /tag= b
FT /product= "NysDI protein"
FT 2806..6906
FT /tag= c
FT /product= "NysA protein"
FT 6952..16530
FT /tag= d
FT /product= "NysB protein"
FT 16550..49840
FT /tag= e
FT /product= "NysC protein"
FT 50260..51015
FT /tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
FT /tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /tag= m
FT /product= "ORF1 protein"
FT
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAEVRIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
PS

```

CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 44.7%; Score 110; DB 22; Length 125401;
Best Local Similarity 85.4%; Pred. No. 5.8e-20; Indels 0; Gaps 0;
Matches 161; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
DB 29738 CTGCGCGAGAGCCCGCGGACGACGACGCTCTGCTGGCACTGGTCAACCGAGAG 29797
QY 61 GTGGCGGAGTGTCTCGGCGACGCGGACCGGAAAGCCATCGCGCCGACGAGTGTTCGGT 120
DB 29798 GCGCGACGCTCTCGGCGACGCGGACGACGCTCTGCAACCGACCTGGCCCTTCGCG 29857
QY 121 GCACTCGGCTTCGATTCACTACGCGCGTTCGAGTTCGAAACCTGCTGATCAAGGCAACA 180
DB 29858 GACCTGGGCTTCGACTCGCTGACGCGGCTGCACTGGCAACGAGCTCAACCGGAAAC 29917
QY 181 GGAATCGGCTTCCTGCTGCTGCTGCTTCGACACCGCCGACCCCTGCCAAACTCGCGGTA 240
DB 29918 GGCCTACCTGCGCGCCACCTCGTCTTCGACACCGCCGACCCGCGCGCTTCGCGGC 29977
QY 241 CACCTG 246
DB 29978 CACCTG 29983
RESULT 8
AA92302
ID AAA92302 standard; DNA; 31422 BP.
XX
AC AAA92302;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..14646
FT /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT CDS 14824..31422
FT /*tag= b
FT /note= "avermectin aglycon synthase protein"
XX
PN WO200050605-A1.
XX
XX 31-AUG-2000.
XX
XX 23-FEB-2000; 2000WO-JP01041.
XX
XX 24-FEB-1999; 99JP-0046961.
XX
XX (KITA) KITASATO INST.
XX
XX Omura S, Ikeda H;
XX
XX WPI; 2000-565458/52.
XX
XX P-FSDB; AAB23751, AAB23752.
XX
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of

PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use
XX
XX Claim 2; Page 134-203; 314pp; Japanese.
PS
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
Query Match 43.7%; Score 107.4; DB 21; Length 31422;
Best Local Similarity 64.9%; Pred. No. 2.3e-19;
Matches 159; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGGCGGACCAACGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
DB 20089 CTGGCGGCGGACGACACGAAACACGACACACCTCTCGCTCGCTCCGCTCCAC 20148
QY 61 GTGGCGGAGTGTCTCGGCGACGCGGAAAGCCATCGCGCGGACGAGTGTTCGGT 120
DB 20149 ATCGCCCGCTCTCGGCGACACACCTCCGACACATCCCCCGGCGCGCTTCGCG 20208
QY 121 GCACTCGGCTTCGATTCACTACGCGCGTTCGAGTTCCGAAACCTCTGATCAAGGCAACA 180
DB 20209 GACCTCGGCTTCGACTCTCTACCGCGCTGCACTACGACCGGCTCTCCGCGACCC 20268
QY 181 GGAATCGGCTTCCTGCTGCTGCTGCTTCGACACCGGACCCCTGCCAAACTCGCGGTA 240
DB 20269 GGAATCGGCTTCGCGGACGCGCTTCGACACCGGACCCGCGGCTTCGCGCTCCAC 20328
QY 241 CACCT 245
DB 20329 CACCT 20333
RESULT 9
AAH79278
ID AAH79278 standard; DNA; 31422 BP.
XX
AC AAH79278;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..14646
FT /*tag= a
FT /product= "AAG65266"
FT CDS 14824..31422
FT /*tag= b
FT /product= "AAG65267"
XX
PN WO200162939-A1.
XX
XX 30-AUG-2001.
XX
XX 23-FEB-2001; 2001WO-JP01381.
XX

PR 29-OCT-1998; 98US-0106100.
 PR 16-FEB-1999; 99US-0120254.
 XX (KOSA-) KOSAN BIOSCIENCES INC.
 PA
 PI Betlach MC, Shah SK, McDaniel R, Tang L;
 XX
 DR WPI; 2000-365602/31.
 DR P-PSDB; AAY92707, AAY92708, AAY92709.
 XX
 PT Recombinant DNA compound encoding oleandrolide polyketide synthase for
 PT synthesizing polyketides comprising a coding sequence for a domain of a
 PT loading module or any one of extender modules
 XX
 PS Disclosure; Page 14-26; 86pp; English.
 XX
 CC This is part of the Streptococcus antibioticus oleandromycin gene cluster.
 CC The oleandrolide polyketide synthase (PKS), also known as
 CC 8,8a-deoxyoleandrolide synthase, is encoded by three open reading frames
 CC (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
 CC "modular" enzyme, where each ORF encodes 2 extender modules and
 CC the first ORF also encodes the loading module. Each module is composed
 CC of at least a ketosynthase (KS), acyl-transferase (AT) and an
 CC acyl carrier protein (ACP) domain. The oleandrolide PKS loading module
 CC contains an inactivated KS, called KS-Q, where Q is the abbreviation for
 CC glutamine, present instead of the active site cysteine required for
 CC activity. The large multifunctional PKS enzymes catalyze the biosynthesis
 CC of polyketide macrolactones through multistep pathways involving
 CC decarboxylative condensations between acylthioesters followed by cycles
 CC of varying beta-carbon processing activities. The macrolide product of
 CC the PKS, 8,8a-deoxyoleandrolide, is further modified by epoxidation and
 CC glycosylation to yield oleandromycin, an antibacterial polyketide. The
 CC invention concerns an isolated recombinant DNA compound, comprising a
 CC coding sequence for a domain of loading module or any one of extender
 CC modules 1-4 or 1-6, including an oleandrolide PKS operably linked to a
 CC promoter. Also discussed are recombinant oleandrolide PKS in which the
 CC module 1 KS domain is inactivated by deletion or other mutation. In
 CC particular, the inactivation is mediated by a change in the KS domain
 CC that renders it incapable of binding substrate (the KS1-o mutation),
 CC rendered by mutation in the codon for the active site cysteine. The
 CC oleandrolide PKS is useful for synthesizing polyketides, which are useful
 CC as antibiotics and molluscs. Heterologous expression of oleandrolide PKS
 CC in host cells such as Streptomyces coelicolor and S. lividans is also
 CC made possible. Unmodified oleandrolide compounds can be provided to
 CC cultures of Saccharopolyspora erythraea and converted to the
 CC corresponding derivatives of erythromycins A-D.
 XX
 SQ Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;
 Query Match 44.8%; Score 110.2; DB 21; Length 50937;
 Best Local Similarity 66.8%; Pred. No. 4.5e-20;
 Matches 157; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 QY 12 GACGTGCGCGACCAACGGGCGGCATTGTCGAGTCGTGACGAGACCATGTGGCGGCACT 71
 Db 29150 GGCGGCACCGGACCCACGCTGTCGTCGCGACCTGTCGCGGCGCACGTCGCCCTGCT 29209
 QY 72 GCTCCGCGACGGGACCCGAAAGCCATCGGCCGCGACGATGCTTCGTCGACTCGGCTT 131
 Db 29210 GCTCGGACACAGTGGCGCCCGGAGCCCATCGACGCGCAACAGGCTTCCGGGACATCGGTTT 29269
 QY 132 CGATTCACTCACGGGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCT 191
 Db 29270 CGACTCCCTGACCGGACGTGCACTAGAAACCGGCTGTAACGCGGACCGGCTCCGCTT 29329
 QY 192 TCTGTCTCGTGGTCTTCGACACCCGACCCCTGCCAAACTCGCGTACACCTG 246
 Db 29330 GCCCGGACGCTGCTTCTGACTACCCCAACCGAGCGGCTCGCGATCACCTG 29384
 RESULT 7
 AAD17186
 ID AAD17186 standard; DNA; 125401 BP.

XX AAD17186;
 XX
 XX 29-NOV-2001 (first entry)
 XX
 DE Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; ds.
 XX
 OS Streptomyces noursei.
 XX
 FH Key Location/Qualifiers
 FH CDS 6337..34771
 FT /*tag= a
 FT /product= "NysI complete protein"
 FT 34792..51099
 FT /*tag= b
 FT /product= "NysJ protein"
 FT 51155..57355
 FT /*tag= c
 FT /product= "NysK protein"
 FT 57503..58687
 FT /*tag= d
 FT /product= "NysL protein"
 FT complement (58786..58980)
 FT /*tag= e
 FT /product= "NysM protein"
 FT /note= "CDS does not include start codon"
 FT complement (59045..60241)
 FT /*tag= f
 FT /product= "NysN protein"
 FT /note= "CDS does not include start codon"
 FT complement (60238..61296)
 FT /*tag= g
 FT /product= "NysD complete protein"
 FT 120628..121308
 FT /*tag= h
 FT /product= "NysR4 (long) protein"
 XX
 PN WO200159126-A2.
 XX
 XX 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001WO-GB00509.
 XX
 XX 08-FEB-2000; 2000GB-0002840.
 XX 10-APR-2000; 2000GB-0008786.
 XX 14-APR-2000; 2000GB-0009387.
 PR
 XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNTP) SINTEF STIFTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE/) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVRIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 DR WPI; 2001-557614/62.
 DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX
 PS Claim 1; Page 188-254; 266pp; English.
 XX

```
RESULT 5
AAF24892 ID AAF24892 standard; DNA; 20394 BP.
XX AC AAF24892;
XX DT 20-APR-2001 (first entry)
XX DE Pimaricin biosynthesis associated polyketide synthase gene.
XX DE Polyketide synthase; oxidative modification; metabolite; antibiotic;
XX KW anticancer; pimaricin; ss.
XX OS Streptomyces natalensis.
XX ST
XX Key Location/Qualifiers
XX CDS 1..20394
XX FT /tag= a
XX FT /product= "polyketide synthase"
XX PN WO200077222-A1.
XX ST
XX PD 21-DEC-2000.
XX PF 14-JUN-2000; 2000WO-EP06227.
XX PR 14-JUN-1999; 99EP-0201893.
XX ST (STAM ) DSM NV.
XX PA Martin JF, Aparicio JF, Colina AJ;
XX PI WPI; 2001-080693/09.
XX DR P-PSDB; AAB31558.
XX ST
XX PT New polynucleotides encoding enzymes involved in the biosynthesis of
XX PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX PT the synthesis of new compounds -
XX ST
XX PS Disclosure; Page 53-80; 116pp; English.
XX ST
XX CC The present sequence encodes a polyketide synthase which is associated
XX CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
XX CC is useful for the oxidative modification of a methyl group of a suitable
XX CC compound, e.g. a bioactive compound including a secondary metabolite,
XX CC antibiotics and anticancer agents. Recombinant cells comprising the
XX CC gene are useful for the production of pimaricin. The polyketide synthase
XX CC polynucleotide may be over expressed in Streptomyces, leading to an
XX CC increase in the biosynthesis of pimaricin, as a source of primers for
XX CC amplification reaction and as probes.
XX ST
XX SQ Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
XX ST
Query Match 45.6%; Score 112.2; DB 22; Length 20394;
Best Local Similarity 66.1%; Pred. No. 1.1e-20;
Matches 162; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTGCGCGACCAACGGCGCGCATTTGCTCGAGCTGCTACGAGACCAT 60
DB 14677 CTTGCCGGGTGTCTACGCGGAGCGCGGAGCGCTGCTCGATGCTGCTACGCGGCCAG 14736
QY 61 GTGGCGGACGTCTCCGGCAGCGGACCGAAGCCATCGGCCGACGATGCTTCCGT 120
DB 14737 ATCGCGACGGTCTTGGGCCACCGCGCCCGAAGCATCGCCCTGACCGGGCTTCCAG 14796
QY 121 GCATCTCGGCTTCGATTCACCTCACGCGCGTTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
DB 14797 GACCTCGGCTCTGACTCTCTGACGCGCATGCACTCGCTAACTGCTCGGCAAGGCCACC 14856
QY 181 GGACTCCGCTTCTGCTGCTGGTCTTCGACCAACCGGACCCCTGCAACACTGCGCGTA 240
DB 14857 GGGCTGCGGTCTCCGGCAAGACCGTGTTCGACTTACCCGACGGTGGATGCTTGGCGGCC 14916
```

```
QY 241 CACCT 245
DB 14917 CACCT 14921
RESULT 6
AAA09469 ID AAA09469 standard; DNA; 50937 BP.
XX AC AAA09469;
XX DT 29-AUG-2000 (first entry)
XX DE Streptococcus oleandomycin gene cluster.
XX ST
XX KW oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
XX KW PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
XX KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
XX KW macrolactone; antibiotic; motilide; erythromycin; ss.
XX OS Streptococcus antibioticus.
XX ST
XX FH Key Location/Qualifiers
XX CDS 152..1426
XX FT /tag= a
XX FT /label= oleI
XX FT complement (1528..2637)
XX FT /tag= b
XX FT /label= oleN2
XX FT complement (2658..4967)
XX FT /tag= c
XX FT /label= oleR
XX FT 5772..18224
XX FT /tag= d
XX FT /label= ORF1
XX FT /product= 8,8a-deoxyoleandolide synthase_1
XX FT 18267..29717
XX FT /tag= e
XX FT /label= ORF2
XX FT /product= 8,8a-deoxyoleandolide synthase_2
XX FT 29787..40346
XX FT /tag= f
XX FT /label= ORF3
XX FT /product= 8,8a-deoxyoleandolide synthase_3
XX FT 40625..41830
XX FT /tag= g
XX FT /label= oleP1
XX FT 41878..43158
XX FT /tag= h
XX FT /label= oleG1
XX FT 43163..44443
XX FT /tag= i
XX FT /label= oleG2
XX FT 44433..45173
XX FT /tag= j
XX FT /label= oleM1
XX FT 45251..46411
XX FT /tag= k
XX FT /label= oleY
XX FT 46491..47714
XX FT /tag= l
XX FT /label= oleP
XX FT complement (47808..49517)
XX FT /tag= m
XX FT /label= oleB
XX ST
XX PN WO200026349-A2.
XX ST
XX PD 11-MAY-2000.
XX ST
XX PF 22-OCT-1999; 99WO-US24478.
XX ST
```

FT CDS 11971...30690
 FT /*tag= b
 FT /note= "avermectin aglycon synthase protein"
 XX
 PN WO2000050605-A1.
 XX
 PD 31-AUG-2000.
 XX
 XX 23-FEB-2000; 2000WO-JP01041.
 XX
 XX 24-FEB-1999; 99JP-0046961.
 XX
 XX (KITA) KITASATO INST.
 XX
 XX Omura S, Ikeda H;
 XX
 XX WPI; 2000-565458/52.
 DR
 DR P-PSDB; AAB23749, AAB23750.
 XX
 XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use
 PT
 XX
 XX Claim 2; Page 66-134; 314pp; Japanese.
 PS
 XX The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified
 CC forms of avermectin and of the intermediates in its biosynthesis, for
 CC use as drugs, veterinary drugs and agrochemicals.
 XX
 XX
 SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
 Query Match 100.0%; Score 246; DB 21; Length 30690;
 Best Local Similarity 100.0%; Pred. No. 1.3e-56;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
 Db 5935 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 5994
 QY 61 GTGGCGGCGAGTGTCTCGGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 120
 Db 5995 GTGGCGGCGAGTGTCTCGGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 6054
 QY 121 GCACCTCGGCTTCGATTTCACCTACGGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
 Db 6055 GCACCTCGGCTTCGATTTCACCTACGGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114
 QY 181 GGACTCCGCTTCCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 Db 6115 GGACTCCGCTTCCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QY 241 CACCTG 246
 Db 6175 CACCTG 6180
 RESULT 4
 AAH79277
 ID AAH79277 standard; DNA; 30690 BP.
 XX
 XX AAH79277;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 XX Streptomyces avermitilis coding sequences SEQ ID NO: 1.

XX
 KW Avermectin aglycone synthase; AAS; avermectin derivative;
 KW drug production; veterinary drug; pesticide; ds.
 XX
 OS Streptomyces avermitilis.
 XX
 XX Key Location/Qualifiers
 CDS 1..11919
 FT /*tag= a
 FT /product= "AAG65264"
 FT /partial
 FT 11971..30690
 FT /*tag= b
 FT /product= "AAG65265"
 XX
 XX WO200162939-A1.
 XX
 XX 30-AUG-2001.
 XX
 XX 23-FEB-2001; 2001WO-JP01381.
 XX
 XX 24-FEB-2000; 2000JP-0047405.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA (KITA) KITASATO INST.
 XX
 XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX
 XX WPI; 2001-582053/65.
 DR P-PSDB; AAG65264, AAG65265.
 DR
 XX New modified avermectin aglycone synthase derived from Streptomyces
 PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 XX Example 2; Page 58-123; 257pp; Japanese.
 XX
 XX The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermitilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is a fragment of the S.
 CC avermitilis genome.
 XX
 SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
 Query Match 100.0%; Score 246; DB 22; Length 30690;
 Best Local Similarity 100.0%; Pred. No. 1.3e-56;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
 Db 5935 CTGGCCCATCAGACGTCGGCCGACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 5994
 QY 61 GTGGCGGCGAGTGTCTCGGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 120
 Db 5995 GTGGCGGCGAGTGTCTCGGCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 6054
 QY 121 GCACCTCGGCTTCGATTTCACCTACGGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
 Db 6055 GCACCTCGGCTTCGATTTCACCTACGGCGCTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114
 QY 181 GGACTCCGCTTCCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 Db 6115 GGACTCCGCTTCCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
 QY 241 CACCTG 246
 Db 6175 CACCTG 6180

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI: 2001-582053/65.
DR P-PSDB; AAG55268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
XX Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the *S. avermectilis* genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;
Query Match 100.0%; Score 246; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 1.1e-56;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGATTGCTCGAGCTCGTACGAGACCAT 60
Db 5935 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGATTGCTCGAGCTCGTACGAGACCAT 5994
QY 61 GTGGCGGCGAGTGTCTCGGCGACGCGGACCCGAAAGCCATCGCGCCGACGATCGTTCCGT 120
Db 5995 GTGGCGGCGAGTGTCTCGGCGACGCGGACCCGAAAGCCATCGCGCCGACGATCGTTCCGT 6054
QY 121 GCATCGGCTTCGATTCACTACGCGCGTTCGAGTTCGAAACCTGCTGATCAAGGCAACA 180
Db 6055 GCATCGGCTTCGATTCACTACGCGCGTTCGAGTTCGAAACCTGCTGATCAAGGCAACA 6114
QY 181 GGACTCGGCTTCCTGCTCGCTGCTGCTTCGACCCGACCCCTGCCAACTCGCGCGTA 240
Db 6115 GGACTCGGCTTCCTGCTCGCTGCTGCTTCGACCCGACCCCTGCCAACTCGCGCGTA 6174
QY 241 CACCTG 246
Db 6175 CACCTG 6180

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
XX AAZ58381;
XX
XX 23-MAY-2000 (first entry)
DT
DE Streptomyces avermectilis avermectin polyketide synthase modules 1+2.
XX
XX Polyketide synthase; avermectin; insecticide; ss.
XX Streptomyces avermectilis.
XX WO200001827-A2.
XX
XX 13-JAN-2000.
PD
XX 06-JUL-1999; 99WO-GB02158.
XX
XX 06-JUL-1998; 98GB-0014622.
XX
XX (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PF12) PFIZER INC.
XX

PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX WPI: 2000-182117/16.
DR
XX Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
XX Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermectilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 100.0%; Score 246; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1.1e-56;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGATTGCTCGAGCTCGTACGAGACCAT 60
Db 6752 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGATTGCTCGAGCTCGTACGAGACCAT 6811
QY 61 GTGGCGGCGAGTGTCTCGGCGACGCGGACCCGAAAGCCATCGCGCCGACGATCGTTCCGT 120
Db 6812 GTGGCGGCGAGTGTCTCGGCGACGCGGACCCGAAAGCCATCGCGCCGACGATCGTTCCGT 6871
QY 121 GCATCGGCTTCGATTCACTACGCGCGTTCGAGTTCGAAACCTGCTGATCAAGGCAACA 180
Db 6872 GCATCGGCTTCGATTCACTACGCGCGTTCGAGTTCGAAACCTGCTGATCAAGGCAACA 6931
QY 181 GGACTCGGCTTCCTGCTCGCTGCTGCTTCGACCCGACCCCTGCCAACTCGCGCGTA 240
Db 6932 GGACTCGGCTTCCTGCTCGCTGCTGCTTCGACCCGACCCCTGCCAACTCGCGCGTA 6991
QY 241 CACCTG 246
Db 6992 CACCTG 6997
RESULT 3
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
XX AAA92301;
XX
XX 10-JAN-2001 (first entry)
DT
DE S. avermectilis avermectin aglycon synthase DNA aveA1 SEQ ID NO:1.
XX
XX Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
XX Streptomyces avermectilis.
OS
XX Key Location/Qualifiers
FT 1..11919
CDS /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 48.936 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180

Perfect score: 246

Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgcgtacacctg 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	100.0	11916	22	AAH79279
2	246	100.0	12381	21	AAZ58381
3	246	100.0	30690	21	AAH92301
4	246	100.0	30690	22	AAH79277
5	112.2	45.6	20394	22	AAF24892
6	110.2	44.8	50937	21	AAH09469
7	110	44.7	125401	22	AAH17186
8	107.4	43.7	31422	21	AAH92302
9	107.4	43.7	31422	22	AAH79278

10	105.2	42.8	65140	22	AAH17184
11	103.6	42.1	1681	24	AAH18438
12	103.6	42.1	11220	21	AAZ87298
13	103.6	42.1	36778	21	AAZ87318
14	103.6	42.1	37948	21	AAZ87285
15	103.6	42.1	38506	21	AAH75633
16	103.6	42.1	38506	21	AAZ56001
17	102	41.5	43280	18	AAH0413
18	102	41.5	53789	19	AAH21187
19	101	41.1	75336	21	AAH14651
20	100.6	40.9	44377	18	AAH78508
21	100.6	40.9	44377	18	AAH80414
22	100.4	40.8	2700	22	AAH17193
23	100.4	40.8	9513	22	AAH88337
24	100.4	40.8	50000	22	AAH88313
25	100.4	40.8	50000	22	AAH88316
26	100.4	40.8	80161	20	AAZ21501
27	98.8	40.2	5676	19	AAH21186
28	97.8	40.0	4689	21	AAZ87299
29	97.8	39.8	1434	24	AAH18434
30	97.6	39.7	77536	21	AAH14651
31	96.8	39.3	6459	22	AAH88336
32	96.2	39.1	15872	18	AAH8715
33	96.2	39.1	15872	21	AAZ87283
34	95.6	38.9	27541	22	AAH17185
35	94.6	38.5	13842	21	AAZ87297
36	94.6	38.5	29879	14	AAQ46806
37	92.8	37.7	561	22	AAH89135
38	92.4	37.6	16767	22	AAH88339
39	92.4	37.6	25360	22	AAH88314
40	92.4	37.6	29736	22	AAH88317
41	89	36.2	13987	18	AAH0415
42	88	35.8	47981	22	AAH30757
43	87.8	35.7	14775	22	AAH88338
44	86	35.0	4209	22	AAH51966
45	86	35.0	4209	22	AAH31625

ALIGNMENTS

RESULT 1

AAH79279

ID AAH79279 standard; DNA; 11916 BP.

XX

AC AAH79279;

XX

DT 04-DEC-2001 (first entry)

XX

DE

XX

XX

KW

KW

XX

OS

XX

FH

FT

FT

FT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.

Avermectin aglycone synthase; AAS; avermectin derivative;

drug production; veterinary drug; pesticide; ds.

Synthetic.

Key Location/Qualifiers

CDS 1..11916

/*tag= a

/product= "AAG65268"

/partial

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

THIS PAGE BLANK (USPTO)

AGAEPFLDGHYORLDADGFGYGVFPQGLURLAARWGDDVYVAEVLPEAGRSDBAEAFG
 HPALDRAALHAAPFTGLSBERGGVLPSWEGVYSHAGATMTTURLVLTVPADVALAGV
 ADGTCAVLSVLSLSTAOQLDTAAARADALFTMTTVPQDT2PCPVALLGA
 DPGLLTHAGDAFPAVPLDLAAAGADGPTVVVLSLWAGTGDAAADPSAHRKAAE
 ALAAQVTLMDHHERFAARLIVFTGATGVPTVAAVGLVLRSAQSGNPGCNFALVD
 DPDCAGVGAALVAALVSEGPQLARGDVLVRVALVRRLRETEVGAGSDGTDGVDGSG
 VYSGEGAVLVTGGLGAVLHARLIVAEYGVRLDILLVYRSGERGAAGELSAELAGV
 GARVYVADTDRAAWELVAGVHA VAAHAGVLDLDMGVWGLTGERLSALPKPV
 AVMLHHEATRGLEDLDAVFDLRLVAGFSGPQSVAAANAFLDAMLRRAEGLPGLS
 LAWNGFSEOSGTMGTLTVFADRLARLSAGPPLSVQAQGLAFDAVAAGTDTATCPVRLD
 LPLVLRARGEVPLLRSLIVRRARRAAVAGSATAGNLAQRLRLDEDOEDMWLDLVRG
 QVALVLGHATGDAGRAPLDFDPSLTAVELNRRLNVTGLRSLPATLVLPVTPVYVR
 LVLVTLDELIGTDAEVAQVPAAVADDPPIVGMACRYPGGVSSLPDLRWLVYRST
 DAVSGPTTNGSNDVESLHPDPDHPGTSTSGSGFHBAGFDEPFMGFSPREALAT
 SOORLLSREAGWELBRAGIDPVLRSRGTVPFAGVDMYSAGMILASPEFSGFOGSGS
 PSLASGRVAVLGLGCPVACTADCSLIVAMHWAQALSGEGCLALAGGVYMSPT
 AVFVDFARQGLSPDGRCKVADADAVGVNSEGSGVVLVLRQDARVNRGHLILVAVRQ
 SAVNGDASNGLTAPOGVPQOQVIRQALASGSLTAGDVVYEAHGTGLDIFIAEQA
 LLATYGRDRPERPLLLQSGVKNGLHTQAAQVAGFKNMILHAGVYVPTLRHAPAS
 SHVDSSGAEVLLSQAOWPETHGVRRAVAGSVSGISNTNHHVYEQAPAKATAAAGA
 ARTPTGAVPVLSSRGSALRGQARLLGHQARDAELDVALLSLTSTSEFEORAA
 VVAQDRQLLASLGAALADRPDAVVEGSEAGRGRTAVLFTGQSQORAAAMRELHEVQ
 PEFAANLDAVCAVFDLPPRLREVVEGSEAGREALLTGTWTPALFAVFAVALFRP
 VSEWGVPRFVAGHSITAAHAAVGLTDEACRLVAARALTMOALTGCAMALTOA
 TEDEIAAHLDDTVIAANGVSPVWISGDEAAETIATFAERGRKTKLURVSHGATPS
 PRMDMLDAFRIVAGELTVAPRIPLVSLDTCRRADAADVCTAEVYVRHVREAVRAD
 CVRTLDPAGATFTLELGSGLLTMAEDTLGDHDHAEVLPMRLAGRAEELAAATALAR
 LQVRGVUDWAAVYLAGTGARTDLPITYAQHAYQLPTPAALAAEDAQDQWLAA
 VERGDAEALDILGSEDQDLTDLPLLALTSGRGNQKHELLDLYRVEVTRLSKP
 TAPVLDITWLLVAGSDQDALDLAGLWRRQVGRHLLDLSDCRYAARLAEER
 AVSTNDPAQVTHPVQAAANGRGVALLPHELRWGLVDLPQVDERAGORLAGLVAK
 DADGEDOVALRATVSGRRLVRHVTVEALPTAAEFTATGVLTGTGGLGAEVARNL
 ARAGQHLVTSRRGPDAPGAELRAELEGKGVSVSVACVDADRLADLRELFVLGFI
 LPTGVWTHAGVHGELPTLSTAFAGLTAKAGAHADLADLRELFVLFQSI
 AGVWGSGNQSA GAANA YLDALALHRRARGLAATSVAWGPWAEAGAADESITLR
 QGLGLDAPMTELRARVRODVTVADVDWORVAPLFTSPASPSALIAGLEPVRAL
 AADERTQDATGASEVVTVALAEPEOLRLITDLATESNTLVGHSSADAVEPEGAR
 RDVGFTSLTAVELRAGLGLSLPSTWFDVTPTLLELAQYLRLELHGVLEHVPD
 ADGADDEPIALIGMACRPFVSGPEQMLDVASGDTLASFPVNRNGWHTGLHFPD
 PDRPTVYTTQGGFHEADFPTGTSIPREALVMPQORLLLETTFESSFERAGIRP
 ETLRSTLTGTFSSGYTGLGAGGTGHEWMTGSSVSLGRLLSVFVLEGFAVTD
 TACSSLLVLAHQSLRNGESLNAVAGATIMTTPNPTAFSRORALAKDGRCKAFS
 DDAGDMLTAEHGVCLVLESAQNGHPLAVLRGSLINQDASNGLTAPOGSPQOR
 VIRQALANLARLPEDIDALEHGTGTPGDPTEAQAQVLTGRDRDPSJALLGVKS
 NIGHTSOAGIASVIRKXNVALHRESLPTDASHPSHWSQVTRLLTQARAWPT
 GRPRRAASVFGSITGNTHVLRLEQAPVADTPAERPAVAPVPTAAVPMVWPTVARGAA
 ALADQGERILHAAETVGTALPAAGPLDTGLSLVSARAFEHRAVFAVPACTDPLAALR
 AVATGDSPPVWAGVADVERGTVFVFGSGSGWQMSGLDSDSAVFAEIRACAAAL
 AETDWSLVDVLRGVGRSLSERVDVQDPAFANVLSLALWRSRGVLPDVAUVHQSQ
 EITAAVYKSLSRDGRVVALRQALIRALAGLQAGMMVSLVLEPLRVEFEGR
 SVAAVNGRSPVWAGPEALDARLHATDIDRARTTAVDYASHSHQVDEHLEELYE
 LAELARLTPSMVQFETITGMDLTVADMAGVYTNLRGRVRFADAVADLAAEYRAV
 EVSSHPLSVAQVTAAGVADVTAAGTGLVLRQDQCTGRTDRLLSAARVFRVGDVWDAGL
 FECTGASRLDLPYAFQHEHLNAVPPAEVAADADPDAAFWTAVEDGVSALTAAGL
 IDEDSVAAIPALTSWRRARRDRSDAWRTRVAKPLGGTLPHPLSPITGLWLLTADG
 IDTIDVAGALETVGAERLDRDECDVRAVRLREAGAEDGTVLSVLAARELTDV
 PGTSLVGLTALVTALQALDAEIDAPVWMLTRGASVSGRDELTAPVQAQVTIGWT
 AALHEPRWCGMTGLPAAUDARAAORLAALVLSGALSDDOLA2TPSGFTVTRVAAE

Query Match
44.7%; Score 110; DB 1; Length 123580;

Query match 44.7%; score 110; DB 1;
Best Local Similarity 65.4%; Pred. No. 8.1e-14;

Best Local Similarity 65.4%; Pred. NO. 8.1E-14;
Matches 161; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTGGCCGACCAACGGGCCGATTGCTCGAGCTCGTACGAGACCAT 60

D_b 29330 CTGGCCGAGCAGCCCGCGCGGACCAAGACGCCATCTGCTGGCACTGGTCAACGAGAAG 29389

61 GTGGGGGAGTGCTCCGGGACGGGACCCGGAAGGCAATGGGCCCCGACCACTGCTTCCCT 120

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Db 29450 GACCTGGCTTCGACTCGCTGACCGGGTGCACCTGCGCAACGAGCTCACCGGGAACCC 29509

Qy 181 GGACTCGCGCTTCTCTGCTCGGTGGTCTTCGACCAACCCGACCCCTGCCAAACTCGCCGTA 240

Db 29510 GGCCTCACCTGCGCCGACCCCTCGTCTTCGACCAACCCCAACCGGGCGCCCTCGCGCC 29569

Qy 241 CACCTG 246

Db 29570 CACCTG 29575

Qy 181 GGACTCCGCCTTCTGTCTCGTGGTCTTCGACCAACCGACCCCTGCCAAACTCGCCGTA 240

Db 29510 GGCCTCACCCCTGCCCGCCACCCCTCGTCTTCGACCACCCCAACCCGGCGCCCTCGCCGCC 29569

Qy 241 CACCTG 246

Db 29570 CACCTG 29575

Search completed: June 17, 2003, 17:55:36

Job time : 479.645 secs


```

/note="Region: acyltransferase domain AT-L"
3485..3724
/gene="oleAI"
/note="Region: acyl carrier protein domain ACP-L"
3806..5056
/gene="oleAI"
/note="Region: ketoacyl synthase domain KS1"
5363..6400
/gene="oleAI"
/note="Region: acyltransferase domain AT1"
7109..7669
/gene="oleAI"
/note="Region: ketoacyl reductase domain KR1"
7943..8200
/gene="oleAI"
/note="Region: acyl carrier protein domain ACPI"
8258..9541
/gene="oleAI"
/note="Region: ketoacyl synthase domain KS2"
9848..10882
/gene="oleAI"
/note="Region: acyltransferase domain AT2"
11723..12274
/gene="oleAI"
/note="Region: ketoacyl reductase domain KR2"
12560..12817
/gene="oleAI"
/note="Region: acyl carrier protein domain ACP2"
13118..24568
/gene="oleAI"
/note="Region: acyl carrier protein domain ACP2"
13118..24568
/gene="oleAI"
/note="polyketide synthase"
/codon.start=1
/transl.table=11
/product="8,8a-deoxyoleandolide synthase 2"
/protein_id="AAF82409.1"
/db_xref="GI:9049536"
/translation="MTNDEKIVEYLRKATVLRKARHRIWELEDEPIAITSMACHFPG
GIESPOLWELLSAGSEVLSFDDRRGMDLDEIVHPDEHSGTSYVRHGFLDHTOP
DTEFGISPREALANDPQORLLLETSLQFLERAGVDPTLKGSTGFGVGAHNGYAD
KVTTPAEAGELITGNASAVSGRI SYTFLEGPVATVDTACSSSLVALHLVAQALR
RGESLAVGVAVMSPDKVEFSRQGLRAGDSKFAASADDFGAEVLSLLLE
RLSDARLHVLAVRGSVQDQASNGLAAPNGSQQRVIRAAADAGLADPADVDV
VEAHGTGRLDPTAQAALLATYCGTSGRPVWLSVKSNIHGTQAAAGVAGVMKV
LALRGVYKTLHVDSPHVDSTGAVELLTEERPWEPEARLRRAAGISAFVSGTN
AHVIEAPAEPEPEPTGTRVAAAGLVVWVSGDAGRLAQAARLAHVSSTGA
GVVDVGSVLAIRSVFBEHRAVMVGTDLDSMAGSLAGFAAGVGVGVGPAEGRV
VFVPGGSGQWVGMAAGLLDACPVFAEAVACAAVLDPDTCWSLVIRGAEVLRV
DVQPALWAVMSLARTWRYGVPEPAVAVHSGQCEIAAACVAGGLSLADGARVVLGRV
RAIARIAGGGMVSVSLPAGRVRTMLDTYGGRLSVAAVNGPSSVTVSGDAQDLILA
GCEREGVTRNLRVQAEVGLVAGMGAFVCSPPHVLPGIEGTLDITVEADAVA
ALDAGYVTRNLRVQAEVGLVAGMGAFVCSPPHVLPGIEGTLDITVEADAVA
LGSURDEGGLRFLTSLAEAFVQGVFVDMVSRTPEGASPRVDTLPYFQRFPLEG
SPALSSNGVEADAFWDAVERDSAVAEELGIDAKALHMTLPALSSWRREQRQR
KVQRWRYVEKWLPNFRAOESLOGGMLLVPOGRAGDVRTQSVAAVKAAGBEATVL
EVDAHLDPRAAALTRWPGVGVSVFSAWEEOALAEHPVLSAGLAALQAQALID
VGGSGESAPLWLTAAVAVIGADTGNVDPVHAQWGFGRVLAALSHPELWGLIDL
PAVAGERGSIIDHAHALLATVLAATWQAARGEDQVARTITGVVPRVRSAGSAR
GARWQPRDVLTVTGMGPLTAHIVRWLNDAGDQVVLGGQAGDGAELRAEFDSH
TTKLEADVTEDSADRLSLDTRGHPLRAVTHAFTVVEFASVABSDLVRFARTIS
SKIAGVQLDELSLSELFGVSGAPLDSVVRNADDEDDPVIIVGMACRPFEGGD
TTLVFDYPTNGLSDVLTADHVPFSSVAGVWGSAGSAAVAGNAFADVAQHRRLRG
LPQTSVAVTPWDDDRSLASIGDSVLDRRGLRALSIPGALASLOEVLDDQDEVHAVADV
DWERFYAGFSNVRTSFDDVHDHRRPALSTAATNDQARDEGGTELVRRLRLTET
EQRELVSLOVEAAVLGHSSTDAVQFRAFREIGFSDLSLTAQLNRNLATTGMRLP
TTLVFDYPTNGLSDVLTADHVPFSSVAGVWGSAGSAAVAGNAFADVAQHRRLRG
TPEAFWLKLEAGGDI SELPANRWDMERLINPDEAKGT SATRYGGFLYDAGEFOAA
FFGISPREALANDPQORLLLETSLQFLERAGVDPTLKGSTGFGVGAHNGYAD
VGGSGESAPLWLTAAVAVIGADTGNVDPVHAQWGFGRVLAALSHPELWGLIDL
CSLAAGGVTVMTSPDSFVSFQOGLSEDRCKAFASTADGFLAGVLSALVLRGE
NSGDLRGYQGVGNAGSVGRVAYSIGLGPVAVTVDTACSSSLVALHLVAQALRGE
DARLGRHVLAVRGSVQDQASNGLAAPNGSQQRVIRAAADAGLADPADVDVVEA
HGTGTRLGDPTIEAQAALLATYCGTSGRPVWLSVKSNIHGTQAAAGVAGVMKVLA
ERGVPVTKLHVDESPHVDNSAGEVELAVEAVPWSRGRVRRAGVSGFSGTSNNAHV

```

Query Match 44.8%; Score 110.2; DB 1; Length 24568;
 Best Local Similarity 66.8%; Pred. No. 1e-13; Indels 0; Gaps 0;
 Matches 157; Conservative 0; Mismatches 78;

QY 12 GACGTGCGGCGACCAACGCGCGCATTTGCTCGAGCTGTCAGACCATGTGGCGGCGAGT 71
 DB 24001 GCGGCGACCGGACCCACGCTGCTGCGGACCTCGTCCGCGGCGACGTCGCGCTGT 24060
 QY 72 GCTCCGCGACCGGACCGGACCGATCGCGCCCGACGAGTCGTCGTCGTCGTCGTCGTCG 131
 DB 24061 GCTCGGACAGTGGCCCGGAGGCGCATCGACGCGGAAACAGGCTTCGCGGACATCGGTTT 24120
 QY 132 CGATTCACTCAGCGCGCTGCGAGTTCGGAACCTGCTCATCAAGCAACAGGACTCCGCT 191
 DB 24121 CGATCTCCCTGACCGCAGTCGAACCTCAAGAACCGGCTGAAACCGGAGACCGGCTTCCGCTT 24180
 QY 192 TCCTGTCTCGTGTCTTTCGACCAACCGGACCGCTGCCAACTCCGCTGTCACCTG 246
 DB 24181 GCGCGGACGCTGCTGTTCGACTACCCACCGGAGCGCTGCCCGATCACCTG 24235

RESULT 14
 AR159871
 LOCUS AR159871 50937 bp DNA linear PAT 17-OCT-2001
 DEFINITION Sequence 1 from patent US 6251636.
 ACCESSION AR159871
 VERSION AR159871.1 GI:16222696
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 50937)
 AUTHORS Betlach,M.C., Shah,S.Krishnakant., McDaniel,R. and Tang,L.
 TITLE Recombinant oleandolide polyketide synthase
 JOURNAL Patent: US 6251636-A 1 26-JUN-2001;
 FEATURES Location/Qualifiers
 source 1..50937
 BASE COUNT 6672 a 16252 c 19273 g 8740 t
 ORIGIN

Query Match 44.8%; Score 110.2; DB 6; Length 50937;
 Best Local Similarity 66.8%; Pred. No. 8.8e-14; Indels 0; Gaps 0;
 Matches 157; Conservative 0; Mismatches 78;

QY 12 GACGTGCGGCGACCAACGCGCGCATTTGCTCGAGCTGTCAGACCATGTGGCGGCGAGT 71
 DB 29150 GCGGCGACCGGACCCACGCTGCTGCGGACCTCGTCCGCGGCGACGTCGCGCTGT 29209
 QY 72 GCTCCGCGACCGGACCGGACCGATCGCGCCCGACGAGTCGTCGTCGTCGTCGTCGTCG 131
 DB 29210 GCTCGGACAGTGGCCCGGAGGCGCATCGACGCGGAAACAGGCTTCGCGGACATCGGTTT 29269
 QY 132 CGATTCACTCAGCGCGCTGCGAGTTCGGAACCTGCTGATCAAGCAACAGGACTCCGCT 191
 DB 29270 CGATCTCCCTGACCGCAGTCGAACCTCAAGAACCGGCTGAAACCGGAGACCGGCTTCCGCTT 29329
 QY 192 TCCTGTCTCGTGTCTTTCGACCAACCGGACCGCTGCCAACTCCGCTGTCACCTG 246


```

/transl table=11
/product="putative lipoprotein"
/protein_id="BAB69301.1"
/db_xref="GI:15824137"
/translation="MGAALITLCTRSLRLAATAGVGGVLAATVLTGTSNAATSGNAG
GNHPTQARINQAQASDARIITTCQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
TGASIPGTLSTGTSWKNPGRLEATKYQAAEAGDVKGSTGTSATITTVSPANDFI
GHLSPEDSTVGAMPVTFNPDKAISDAQVQSEIRVSSSGSGGSDVDRHDFR
PENYRPGSTVTVTLNRHGIQKTVFKIGRSQISTVDAKTKMTVDRGDKTIPIS
SGSEHTYVQMWISSEKFRQIHNGASVGLTEKNGSDIDIKAVPHAMRLTDSGTI
HGNTVGDTSVGEANTSHGCCVGLDKVRGDSHGROPAAFDHFDSIGVDVIVKNSSEKT
MLAPDNLGSDNWSSEWVAGSPAT"
/complement(1256..2290)
/codon_start=1
/transl table=11
/product="putative transposase"
/protein_id="BAB69302.1"
/db_xref="GI:15824138"
/translation="MPVEFLSHDQVSRYGRFASAPTFGELEBQFRLDAKALASARXR
APANRLGWAQVQGVMLGVPPTEDLSVPEVVRFAAQOLGVDPGFAAYGARRQNR
YEHAWETRAVAYRSPAQAQVREFLATRVWASMEGPRALFDRAGVWLDNRVLLPG
ITTLTRVAVRAQENALYTLDEAVPEKLRQSRDLKVPGEGRVSELERLARLSPK
GVSGADRLAESEIRMEAVNTAGRGSTDIKELRVSVQSVQRWGLSDGRPGRG
AVPQSVSELSSETLFAALDEERAKGPAPSWTIRYLHDFPIFGORYAYKWSQGCLHSF
GRSAQYRDO"
2935..21081
/gene="pteal"
2935..21081
/gene="pteal"
/codon_start=1
/transl table=11
/product="modular polyketide synthase"
/protein_id="BAB69303.1"
/db_xref="GI:15824139"
/translation="MVSELNATNRDCSASDAGIATVIGLCRLPGAADPAAFWR
LLSGADALTVDPDRDGVAVADASAPDRTGIRRGFLDSIGDFGLSGSPKEA
AMDQRLVLEALDEADHVRAATLSTRTGFGVGMAMDDYATLHRRSLTASIH
TVTGLRLIINRVYSFLGNGPSLTVDGSSSLVSHVLAACESRLGTEIALAGV
NLNTVPESALAAKFGGLSPDRCFTDARANGVYRGGGGIIVLKLARAVADGDPV
VCYIRGSVNDGGDGLTVPLSQSQEQVLLAYERAGVNPADVGYVHLGCTGTKEVD
PIEAALQVIGREPGRPLRVGSATNVLHLEGAAGITGLLKAVLSLNHREAPSL
NFTGNPAIPDLRLNLRVQTHSEWAAEDGRPLLAGVSSFCMGCTNCHI VLAGAAVE
AGEATVGGGAAPTFWPLSAKTASGLRAQAALIAHTPEALPLPDVGSLATGRTA
FEHRAVGVVDRDPLRALRELSTGGLDIAALTGRAGRGELALFLFGSQGSORAGMR
ELASVPASALDEVCACHDPLPLRPLREVMFAVGTGEAEIDQTLTCTLSFAVE
VALFRLESSEITDVLNMGHSIGELAAAHVAGVLSLADACALVARGELMQALPAGCA
MIAVQAEDVRIRIGERTDRVSIAALNGPDSVVVSGDEDIATSIAPAAALGRTSR
LRVSHAFSPHMEPLMAEFRNVAEGLTFHAPRIIPVNSVTGLRSESPGYEGCTAAVY
VRHVREAVFADGARLDEQGVRTYLELPGCVGLASMARAGDSIESLFPALTRRPE
AQLAAVASLHVHGVPEMDALPGGRGARRKVALPTYAPERQRYLDGDTTSPVAT
QVETWTFKIDFDSLSVELRNQLEATGLRPSGLLFDHPTPALARHLAAQSLGG
QDAALAPATVADPHEPTAIIVAMSCRLPGGVSSPEDLWRLLASGGDGLSGFTDGRMD
VETLYDPEGTFPGKSTRHGGFLHEAAEFDAAGFFGISPREAAADPOQRLVLETAWEA
FERAGIDPATLRGSRAGVFIGATQDQGRPLRHEPADGVGGLTCTTTSVAGSRVYS
FGLGSPAVTDTACSSILVALHVAQSLRSGECTMALAGVTVMTSCMFMFEFRQRG
LSADCRKCAFAAGADGTGCAEGVGMMLVERLSDARRNGHOVLAVRGSAINQDCASNG
LTANPGSQQRVQALAAAGLA PADIVDAVEAGTGTLLGDPPIEAELIILVAGSGRQD
DRPLWGLSKNIGHTQAAAGVGVKMWAMRHGVLPLTHLDEPTPHVDMWASGVGR
LSTDAVENPQARLLSLDAEGVGGAGVRPEDVALSLATSRVLRRAAVAGGVDLAEIR
QGLBALASGASCAAPVGGVAGVRGFLPSGQGSORIGMRELYAAYPVFAAAYDEVCA
FLDAPVDVDSSELNQTGTTPALFAVEVALFRLESNGIRPDYVAGHSVGRIIAAHVA
GVLSLEDAKLIVSARALMQALPDGGMVAVQATEDEVLRLHTGVGGLIAVNGRSVV
VSGAEDAVTAGEITFAGRKRTSLKVSHPMLDPMLEFAEVVRGLTFSSESRIP
VNSLNTGLBAEYVTPYVWVRHREAVFADGLETIGELGVTFVVEGPGVLSLAQ
CFDDGTDITVTVLRADRPPEYALVTAIGRLTHGIGSPMQVLPFGARRVDLPTAFQ
RERYWLDAPADAMPDAEFWDSSESDRASLGALLGLEPTELDVAASKLSAWR
RKREOSTTDCWRVRLTWQPLDDMVRAAPSGTMYLVVSEETAWTEARLFRAGLTGVAL
VPPVAGETDRLARAPADQAGQADGILFAAQQGEGAGANALRVLVQLAGDAGI
BAPLWCLTSGVSTGTADLLTDPVSAQWLGRIVALEQPOQWGLVDLPABLDSRAL
ERLAGLAGADEQLAVRASGYGRRLVRPLSAPADAPWSPRGTVLVTGTTGALGG
QVARWLAGAAGEHLVLTSSRRGPDAPGAELKAELEELGAQVTVVACDAADRDAELP

```

CDS

gene

CDS

gene

CDS

```

GRYIVNAVHTAGVLDGGLVESLTPERLNHVLRLPKVDAAHLHLHELTRDRQDLDAFVLFP
SSMTGVKNGGQGAYGCAAFDLDAEHRRRAQGVPALAVAMGWSADGMADGAAGDHL
RREGVRAIALPAISVLHGALTHGETSMTVADVDMERFVPAFVGTGRCPLTLCGVPEAR
EALBARAAQSPSEVPSALVQRLLATGTGECQRILLDLVREQAATVLTGHTGAPBA
DTRFRETQDSUTAVELRHLNTATGLRLPTLTVFDHPTTALARHLGEBELGRHGUE
TAAATAAAAADDEPLAIVGMSCLPGGVSPEDLWELSSDRMNSGPTDRGWDID
ALYDPOFGPKTYSREGGFIQAGDFADGLFGISPREALAMPQORLLLETWEAPE
RAGILASVRASRTGVPIGTNGQVYANGRLNRAPEIEGVALTGKAASVVSRTSYTFG
LEGPAVTDTACSSILVALHVAQSLRSGECTMALAGVTVMTTDPLEVFPSRQGLS
DPCRKAPAGADGTGEGVGLLVERLSDARRNGHOVLAVRGSAINQDCASNGILT
APNGSQQRVIRQALASGLTPADVDAVEAHGTGRLGDPPIEAQALLATYVQGHSDDR
PWLWGSKNIGHTQAAAGVGVKMWAMRHGVLPLTHLDEPTPHVDMWASGVGR
TEPQWPGTTPRAAAYSSFGIGTNTAHTIEEAPAAAAPAREHPRVPAWVLSAK
SEVALRAQELRALFATDVS PVDAGFSATTSALREHRAAVTSGPAPELRGLALELA
AGEPAANVAGRAHSAKGVFLPSGQGSORIGMRELYAAYPVFAAAYDEVCAALLDAT
YVDATLHQGTGTQPALFAVEVALFRLESNGIRPDYVAGHSVGRIIAAHVAGVLSL
EDAKLVSARALMQALPDGGMVAVQATEDEVLRLHTGVGGLIAVNGRSVV
DAVTAIAEYFAAQGRKRTSLKVSHPMLDPMLEFAEVVRGLTFSSESRIPVWSNL
TGRLTESYTPYVWVRHREAVFADGVLTHELGVTFVIEIGPGVLSALVQCLDDG
TDITVTIPALRADRPPEQALTVAQLHVGCAEVDMAHFPFGARRVELTYPAFQORVY
LDPASAGDMRAVQGEAGHPILGAAPVLAADSVGLLHGLTGRLSKHTHPMLVDHVASGV
LLPCTAFVELAITAADAQVGLLEELTLETPLAVERGGVALQVHVGADDSGRRLPT
VHSQIDDDGDAIHEDDSERAWVRHSGVLTATGPGVDGSLTAMPVPGABPVDVGY
DRLAGMALDGPVFGQLRAWRSGDDFFFAEVLPGAERTDAGAYGLHPALFDALHTV
WLGAVEPEAGTGNGLLFFAMSGVRLAAAGASDLRVKVSAGTGTSTVSLADGTGTP
VEVDSILTRPVPADOLRAGSGDDAVFGLKMTVPVSLPAPDGMRIETVYEDLAALRAT
DPPADTEADAADAADAAVVPCTGTCADLATRVRETVHVLVWMLAEERPGRLVL
VTGGDLASAVCGLVRSQATNPDRILVLTSSDDAHAGNGSVDADADLAVLPLGLIA
SGEPQAADHADDEVVRPVLVAVPDAEPGLPGSGTVLLTGASGGLGGLPARHLVA
EHGVRSLILVSRGGDAPGAELTADTAQADVTWACDVAADRAVALLAGPQGL
SAIITHTAGVLDGIIIGSLTPELRDAVFRPKVDAALNHLHLAGSGLSALVLFSSVA
GLTGTQCAANYAANTFLDAEHRRAAGORATSLAWGLWAQDSAMTSGLDHTDIT
RIORMGLAIPSSDGLRLFDAALDRAVVALRLDLGLGQTVPPVLRGLRGTVRGT
ARTQATKPGSLQORLAGLAEAEREQVLDLVRTEVAVLHAGAGSACADSPKDI
GFDLSLTAVERNRLNSAVGRMLPATLIDFPNPLTLTLAHLTGBAAGEAATVPTTUS
TGTDADEPTAIVGMACRPGGVSPEDLWQLVFGSDRDAVSGPEGWDEQVLDYN
PDQWGTSTYREGFLHDAADFAEPFGISPREALAMPQORLLLETWEAPEAGIDIP
ATVGRGTVFAGWTVGGRVHTSPGLEGVLNMGSGAGSIASGRVYTFGLGEPV
TVDTCSSILVALHVAQSLRSGECTMALAGVTVMTAGPSVFPVFSRQGLSADGRCK
AFGAGADGTGWAEGVGMMLVERLSDARRNGHKLAVVRGTAVNQDCASNGLTAPNGPS
QQRVIRQALASGLDEVDVAEAGHTGRLGDPPIEAQALLATYVQGHSDDRPLWLSG
LKSNIGHSQAAGGVGVKMWAMRHGVLPLTHLDEPTPHVDMWASGVALLTEAVSW
PASDGRRAAVSFGVSGTNHVIIEQAATSEVTAAGTAVAGVPVWLSAKSADALRE
QAQRLSLFDLDAEGVGGAGVRPEDVALSLATSRVLECRAAVVGEBELGEPQGLGIVS
GAPVGSAGVGRVGLFPGSQGSORIGMRELYAAYPVFAAAYDEVCAALLDATVVDVAE
TLHQSTQPALFAVEVALFRLESNGIRPDYVAGHSVGRIIAAHVAGVLSLEDAKL
VSARALMQALPDGGMVAVQATEDEVLRLHTGVGGLIAVNGRSVLSSEDAVLDI
AEVFAQGRKRTSLKVSHPMLDPMLEFAEVVRGLTFSSESRIPVNSNLTRLAE
VYIPEYVWVRHREAVFADGLHTLHGLGVTFVIEIGPGVLSALVQCLDDGTHITV
PALRADRPPEYALVTAIGRLTHGIGSPMQVLPFGARRVDLPTVAFORERYWLDAPQA
SGDVRAAGLADHPILGAATADSDGVALTGLLSLGTGHPMLADHAVNGIIVLLPGTA
PVELIIRAGDETCCGMIETLLEPLVPERGQVTLQAVGPEDESGRRLSVHSRVD
DRSLRHATVLSLSAEPKAGKAGAMPAGAEAVDLTA FYEGMAAGLDTGVPVQGRUR
SVWRDGDVFAEVLDPESAAAAFALHPALLDAALHALAAGLVLSDSPALPFAM
GQVQVQAGAAAMVVRKLSATGTSVTLAVADGADPVSASVLSRAASQOLPGAVG
SDGLFGLMETFTLPSATDALMETVADFDALRESECTPDVVVVVPCPAGSGRETA
DRVHSTSEVLAVOMWLAEERAGRLVTRPGDLAHAAVGLVRSQAQSNPPDRIVLV
EASETDEAVRVLPAALASCEPQFAVRDGEVFPVRLVKAETSAATGTGTGDPFGVPL
LTGASGGLGVARHLVAEHGVSLLLLSRRAEAPGAELEAELEAANGABRWAACD
VADREAVGMLNGEQSGSNAVVTAGVAGDNGFVSLTPERMREVRFPKVDVAVLNLHE
CTRDMGLAAFPVFSVAGVWAGSQAQYAAANGFLDAFSAHRRERGLPAISLAWGWE
QSGMTDGLVPAADARMAKSVLPLPPEEGLRFDALASDAEVALPVRIDTGTALRAG
EAPVLRALVPAADARMAKSVLPLPPEEGLRFDALASDAEVALPVRIDTGTALRAG
CHASDRVREHAFQDLGPDLSLTAVELRNLNKAATGLRPLATLVFDHPTTALLARHV
MAAARAPALVDSLADLDRVSEQLFTTILASAEKDRILSLKQLLAIATAGEFGKAPA
QEGEPGSDLSATDDEVDLILGKFGFIS"
21273..31847
/gene="pteal2"
21273..31847
/gene="pteal2"
/codon_start=1
/transl table=11
/product="modular polyketide synthase"

```

```
/gene="olmRII"
/658..10486
/gene="olmRII"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69190.1"
/db_xref="GI:15823973"
AGSAAETAEFGVVGOLFPPPTGPGAGFLARATGSDAGPGLDADRLFDALCOLERR
RFTVAVDDADADAASLRFLRYLRLRTAPVMVLTEPTGPDALPSPFOELLRHP
QCRRLGLPLDGTQVVSQVSDRVRPLAADFHALSGNPLLVHALLADHRAHR
LSPVSPFTVPSVSSAASASASASASASAPASAPAFKAAALAWAYRDEFLF
DVACFPAVLGSAATPARVACLVDRGADAVQMTALDTAGLGGTFRSLAKWQLE
HVVETAEELRRRLRHGGAPATVVARHLLAAPVAEPWAGRTLLAAASAKMOEGR
EELSLDLRAGAQATTEQERAAVVMARVKIGWEIDPLVSPMLDELVAALRRGDLS
GDAAWIKHLAQQRPEAVESLSTLAAGDAGBPGVEAEHLFVRHLRGLSCPLEAS
VPGSNPRTPRPRRVSASVAGEVLGKVLTEGPDHHTVMAEELHLCGRYSVTV
EALRAALLAVYAGPRDAFWCDGLLQOAGRTSTGTSVAILSGIRAEIALRQALPA
VEQYARSALSISLPGMGVAVGSLPAAALVHSTTASGMSGVAGAWLNOQVPMFRTFH
GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEMDMVPTFLPWTSSAAOALGN
PGOARALVREOLGRPGSTARGVALRLLAATGEPGRPRPALLRESNCSGDOVE
LRLALDECOALDHVGAAPAKARNTARHARNVADACGGDALFRKLFREBVPDRGDPG
LITGEQASLTAERVRTALAAALGHNSREIGRKLFITKSTVEQHLTRVRKJGVNR
ADLGDLAGTNLGGTANMASSAVAG"
/10933..11271
/codon_start=1
/transl_table=11
/product="p450-like protein"
/protein_id="BAB69191.1"
/db_xref="GI:15823974"
SFGFIHOCISQSLARAEPOIAETLLRLRLPLRLAASFELRFRDGVGVKELPD
PCPALPDRP"
/11458..29898
/gene="olmAI"
/11458..29898
/gene="olmAI"
/codon_start=1
/transl_table=11
/product="modular polyketide synthase"
/protein_id="BAB69192.1"
/db_xref="GI:15823975"
/transl_table=11
/product="modular polyketide synthase"
/protein_id="BAB69192.1"
/db_xref="GI:15823975"
VVDLPLADADGDDHICILAGSANNDDGGEGLTVPDRAGQEAVALTAQEAGTSP
AAVGYELHGTGTPTGDPVEAAAVGAVLGGAGPAGQLVSGVKNTIHLGEGAGTIG
LLKAVLVRHREIPATLHASPNAIPMAELIGVNTALTTPWEPQAGLPVVGVSFGM
GTTCHVLTPEAGYDAAAPATAGIRPAGTPAPLLVSGRDEQALRAQAKLRDLHPT
HPGLPDRVATLAAAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
ALAFVFGQSQRPGMTAEILYRTSAEYAAALDEVCASLDPLRRPLRLDLPFAEGSDE
AALUDRTEFTQPALFVAEVALFRYAEGHCLTPREFLHSGVSGELAAHAGVGLSLADAC
TLVAARQMDRPTAGMVSQVATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
AVAARMAARGKTRKLRVSHAFSPHMDGMLTEPRIAAGLTFCGAPRIPVSNETGVL
LTDADARSPEVYRHAARVAFDLGVLLDQGVTTLELGGPGLVSAMARDCLRDAG
TVCVPLGRDRTPPEALLAALAGLYRGVPMQWERTATAGARRVPLPYAQORRH
WLPEFSASGRPEVPEAGAGQAGQAGQAGHAGHDGEPAPARDRTTPELVR
DRTLELVRDATALQVLHASPDDVATHEAFRELGSMLAEILGEMTEATGRVPTT
LDFDPTDLARLEADGAERTDAGVPSPFAAAHDDDDPIVIMGCRPLPGVRSPE
DLMFVAAAGDALTALPVDGWEPFADFPGEGLPDAVDFAGFGISPREALAMPD
QORLLTSNEALERAGVDALKLRGSRGTGVFGASPEYGRPHLETSDAGHVLGTGA
PSVLGRIAYVLGLPALVTDTACSSSLVALHLLAALRGECDDLALAGVTVMATS
GMFAEPARQGLADPGKAFADGADGTGMEGVGMVAQVRLSDAVDRGRPLVAVRG
SAVNSDAGNGLTAPNGSQOQVIRQALASAGLSAADVMDVMEAGTSTSGDPIEQAQ
LLATYQGDPRDPLTGLWSKNSNIGTQAAAGVAGVTKSVLARNGLLPKTLHVDPT
PEVDSAGAVELLTEGREVETDGPRTGPRYSAFGIGSCTNAHVILEQAPSVPEPTPA
DGLLVPMVVSARSEERALRAQARVGLRPAVGLSLAVTAGLEHRAVLVA
EDEAGFLHLDALDALETPAGVVEGTATLFGSQSQSQRMGRELYSAPHPFAAFD
EVCNAFGLPLGALREVFGAGLDRTEWAQPALFALEVALFELVSSWGVARDVLGH
SIGLEAAVYAGVWSLADACRIVAARGLRMOALPEGGAMVAVEADELDPDLPQGVSV
AAVNGPRSLVSGDEEPTALAQAFAEQGRRTKQLTVSHAFHARMPEMLAEFAETLA
```

RSVVSASETAVTLVQOLSRLAETDRREVLDLVAGQVAALVGHAGTADIDPRRLREL
GFDSLTAVELNRLSAATGLRTAATVIFDHPTVDALAAHLDDGLMGSEAAAAADAPT

Query Match 45.6%; Score 112.2; DB 1; Length 84985;
Best Local Similarity 66.1%; Pred. No. 3e-14;
Matches 162; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 1 CTGCGCCATCAGAGCTCGCGCCGACCAACGCGCCGATGCTCGAGCTCGTACGAGACCAT 60
Db 66476 CTTGCCGGCTGTCCACGCGGAGCGGCGCGAGCGCTGCTCGATGCTACGCGGCCAG 66417

Qy 61 GTGCGGCGAGTCTCGCGACCGCGACCCGGAAGCCATCGCGCCGACCATGCTTCGCT 120
Db 66416 ATCGCGAGCGTCTCGGCGACCGCGCGCGGCGGAACGATCGCCCTGACCGCGCTTCAG 66357

Qy 121 GCATCTCGCTTCGATCTACTCAGCGCGTCCAGTCCGAAACCTGCTGATCAAGGCACA 180
Db 66356 GACCTCGGCGCTCGACTCCTCGCGCGATCGAACTCGCGTACCTGCTCGGCAAGGCCACC 66297

Qy 181 GGACTCGCGCTTCCTGCTCTCGCTGCTTCGACACCGCGCCCTGCGCAAACTCGCGCTA 240
Db 66296 GGGCTCGGCTCCCGCGCAACGCGTGTTCGACTACCGCGGTGGATGCCCTCGCGCC 66237

Qy 241 CACCT 245

Db 66236 CACCT 66232

RESULT 10
AB070940 104326 bp DNA linear BCT 22-JAN-2002
LOCUS Streptomyces avermitilis oligomycin biosynthetic gene cluster.
DEFINITION
ACCESSION AB070940
VERSION AB070940.1 GI:15823967

KEYWORDS Streptomyces avermitilis DNA.
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)

JOURNAL 21477403
MEDLINE 2 (bases 1 to 104326)
REFERENCE Ikeda, H.
AUTHORS Direct Submission

JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
TITLE (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)

FEATURES Location/Qualifiers
1..104326
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
complement(1..1134)
/gene="mcma"
complement(1..1134)
/gene="mcma"
/note="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=11
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
/translation="MTGYSTADANARFQLVGSSTRVAVVDFLPTRAGLSDAPAA
RGLVRRGVAVDSIDMRVLFGLKLDKRVPSLAVDAPAPLLVLYQLVAEEHGAQAQ
RLAGAVRDFKFLARGPGVFLPRLPSRLAMNVPAYCQALPRWQCLSVCGHLAGT
GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCVLGGTTGRENRAKLRAATR

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene

VWTRANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATATLVRIERLGGTLTVL
EHLGAGQDGLLRPHGDRPATPRPHPHQDSDLQARQTERLAKLRWVQPAVDEALLR
LKAACGDDNVLVPMREALAARATVGEVCDALRSTWGTGSPAT"
complement(1131..1886)
/gene="pptA"
complement(1131..1886)
/gene="pptA"
/codon_start=1
/transl_table=11
/product="phosphopantetheinyl transferase"
/protein_id="BAB69186.1"
/db_xref="GI:15823969"
/translation="MRAHEVFADVPVSSLHPREALVQDAVESRRREFATGRHCARRA
LPSLGVPGPLATPRGAPRWPGVCGCAITHCYGAAAVARAPGAVIGDIEADPRP
LPGVLEAVALPAEQFOVKHPTARHPGVVMGKLLFSAKESVYKAWPATGRPLGFEEA
LIDLPVSSFRVRLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVVGGPRDTPHAP
SPDRTKSAPTGSRADRGCRPTDRATDRPTA"
complement(1910..2875)
/gene="simX4 homolog"
complement(1910..2875)
/gene="simX4 homolog"
/codon_start=1
/transl_table=11
/product="simX4 homolog"
/protein_id="BAB69187.1"
/db_xref="GI:15823970"
/translation="MSQDRHDGSTAALGGPSHGRLAVSDIHRVHRQNRDI VQDLHP
GCAEDMLIVAGDAESVADITWALSLLRDRPAKVIWAPGNHFWLTAPGDDELGERK
YRHLVDLCRALGVVTPEDPYVWHGAGGPVAVAPLFLLYDYSFLPGCAKSHALRLA
DDAGVICTDFLLHPDPYPSREACARVATRLARLDLDDPTLVLVNHPFLVREPT
RVLRHPEFALWCGTTATADWHLYRAAAVHGHLLPTIRICDDVPHHVSLSGYPREM
QRYGRPDMPVQILTPAPAPAPVPVPVPPAPPCHRTASPADRRP"

3213..3971
/gene="olmC"
3213..3971
/gene="olmC"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="BAB69188.1"
/db_xref="GI:15823971"
/translation="MSANRVDGNLWVRQHPAPTAATKLVCLPHAGSASVFLPVSKA
LAFQVDVLAIQYGRQDRRHEKIDNIAELADAVTAVLEPADRLPLVFGHSMGATLA
FVALRLERKIGIVPLAFSGRRAPSRHDSVHLRGDDDEIIAELKTLGSTRMSGLD
EELRMVLPAIRSDYKAAETRYAGGARLTAPIHAHFGVEDPRVSLDEAQAQADHTTG
HPLRSHPGGHFYLNDQAPRVEIARVSSVAS"

4643..7438
/gene="olmRI"
4643..7438
/gene="olmRI"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69189.1"
/db_xref="GI:15823972"
/translation="MLRGQEIIDLRAAYSSVRSRAVLVESAGCGCKSBILE
LVHVESIGLIALRATAIERNTPLSVLNVNAPFPDDLVRPRTMDSAAARMIS
TEGRDKARRDAWYELMRTCTEVRALAQAPLVAVDDIHHGDADSLROLLHARHC
RNVPVLTFTQSLAGGPADRTFGAELLRQNLHRLPRFCRQDIADMLAEQGRADAG
QGTVERYFVSGGNPLLLRALLEDLRLEDDTVDQDVGPAGQAYGRAVLACLDGGG
EKAHEVAAGLAVLGGAFTHELLRLDISAALTENLRALRAAGIVVGTGFQHPAAKA
AVLDSLEPSHRLRHORSAALHAGAPALAAHLEAGQADGAWAVGLVQQAEEQA
LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRVRSADAEHRVDDALRGFGA
GLITPSDTGSLAALLAHLEQARSVLEQSRPGCDADRTAARQALAEIRDSVLNR
GAWCAGTRAAHENPAGPSTGRAGTGRALWLSGSHAALLAQRIAGTEGLLEIHTLTD
ATFDLVNAVMTLALGASERALHWCDFLEAKRRDAPGWEAVFAVRAEAGLURGN
LAEARDARRCFVLPYRHSALRGSAVSYLVTQIARGYEDAAARQLSRPMADSLPD
TIHLDYLRRAGLYMARTQYHSALRDFHEVGRLAERWGTDMALLPRTDTABALLR
LGRREQRLATEOLAMVQDGPRIKGISLRLOGLASEPAARLKLNRVAVELSRADD
LRELSRTLFDLADAHRETRQTRONMVLREAROLAKSCGAQPPDRIGSPDAGASGG
LFGEGPERRNADWHLSESERRRVAGLAASGTTNREISVRLYITMSTVEQHLTRVTRKUK
INCREELLARAGVQV"
7658..10486

Ddb 14677 CTTGCCGGCTGTCCACGCGCGAGCGGCAGCGCGTCTGTCGATGTCGTACGGGCCCCAG 14733

QY 61 GTGGCGGCGAGTGCTCCGCGACGCGGACCCGAAAGCCATCGCGCCCGACCAAGTCTGTTCCGT 120

Ddb 14737 ATCGCGACGGTCTTGGGCCACGCGCGGCCGGAACGATCGCCCTCGACCGGCGCTTCCAG 14796

QY 121 GCATTCGGGTTTCGATTACTCAGCGCGCTCGAGTTCGGAACCTGCTGTATCAAGGCAACA 180

Ddb 14797 GACCTCGGCTCGACTCCCTGACGCGCATCGAACTCGGTAACTCTGCTCGGCAAGGCCACC 14856

QY 181 GGACTCGCCCTTCGTTCTCGCTGGTCTTCGACCAACCGACCCCCTGCGMAACTCGCCGTA 240

Ddb 14857 GGGCTCGCGCTCCCGGCAACGACCGTGTTCGACTACCCGACGGTGGATGCCCTGCGCGCC 14916

QY 241 CACCT 245
|||||

Ddb 14917 CACCT 14921
|||||

RESULT 9
SNAP278573/c

LOCUS SNAP278573 84985 bp DNA linear BCT 17-MAY-2001

DEFINITION Streptomyces natalensis pimarinin biosynthetic gene cluster.

ACCESSION AJ278573

VERSION AJ278573.1 GI:12055067

KEYWORDS ABC transporter; cholesterol oxidase; cytochrome P450 monooxygenase; ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine dehydratase; orfX; pimaA gene; pimB gene; pimC gene; pimD gene; pimE gene; pimF gene; pimH gene; pimI gene; pimJ gene; pimK gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene; polyketide synthase; sensory transduction protein; thioesterase.

SOURCE Streptomyces natalensis

ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 84985)
Aparicio,J.F., Pouces,R., Mendes,M.V., Olivera,N. and Martin,J.P.
A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimarinin in Streptomyces natalensis
Chem. Biol. 7 (11), 895-905 (2000)

JOURNAL 20547809
MEDLINE 11094342
PUBMED

REFERENCE 2 (bases 1 to 84985)
Aparicio,J.F.
Direct Submission
Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics, 24006,
Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
SPAIN

FEATURES
source location/Qualifiers
1..84985
/organism="Streptomyces natalensis"
/db_xrefs="taxon:68242"
<1..2086
/notes="ORFX"
/codon_start=2
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC20917.1"
/db_xref="GI:12055068"
/translation="TRPADFRBATPYGALLAPMPERSADRVETKWTSQWQSQRFL
THSDATELVRLVDSGLQHGVPLVLLEDQLQWADTASDLIRLLSPRCQGHPLGI VLTARKFE
ATDVCELRITLTVGRPRPTATLRLCGLPROAVGALVEAHAGSGVDPEVLEVRHSRSE
GNPFYMQLLSLVGDRBLRRPDVAALLAHTPTGVREALNORPAALPEPVLRVLRCLC
AVLGAEVDTDLRLRRATCDASNAGLESNAIRAGLLGEDPHPGQLHPFAHALVOETLFD
ELPRADRNRHAKVADDAFYARRRHGVDEEIERVAHSHWAKSAALPAEAHLPRLLGJAA
ERAEQAQEVSIVWLRRRAHVILGF LPDPDPATVLEQKLHQLQGOVLRTATRGYGHPE
AETALTRGASTVETHAPDDPSVWALCAASLVTCGRYDASROLGSLLRDLSHQTRREY
AVLGAAYGSGI VLHVGRMLPOALEHGVGERADHPANEHALARTFOHDPRVSCRSY
NFETHWILGHROIADARRELISLTEYESRPDSRFALYVDGVAAWEGDFTFLASS
DEGVRVAGDHGLLYWKAMIGVVGMGLVHAGDORDGLAMHTLSLAELGOSRYLRHPI
HLGLLGOAQHNHGQIEARTTTFMILLAAVEVQRERYLHPALPATPLLHELIGHGADE
ALADRDPPGRGV"

CDS

SPGRHARTHPRTGRGLFDAAALAGDAHLLPVRFDOWASLRAQCEVPPLLRGLRITRARE
 SVGSSAAAGLVRLSGRGTVERREVLLDVRQIAVVLVGHANPETESTRETVFQDLG
 FADSLTAVELRNLNNAKTLRSATAVFDPTADALVDVVLGHANPETESTRETVFQDLG
 PAGAADDPVVLVGHSCRYPGVGSGSPEDLWRLVSEGVDAVSDFPTDRGWDMVSLYSDDP
 BALGTSYTRSGGFLHEAAEPDPDPFGMSPREALATDAQORLLLETTWEAIEATGIDLPA
 SLRGSTPFGVAGMYTVDGDLIDQFEGYIRGNSAASIASGRVSTYVTFGFGPAVTD
 TACSSSLVLAHWAQASRSCSLAVAGCTVMSTPTTFBFSGRGLSADGCKKCAFA
 DAADCGWEGVGLVLSRLSDARRNHRVLAVRGSAVQDQAGSNQITAPNGPAQOR
 VIOALASAGLSAADVAEHAHTGTGLDPTIEACALLATYGOBRPDRPLLGLGSVK
 NIGHQAASGVAGYIKMVLAMRHGVLPRTLHVDEPSHVMDMSGAVELLITSEAEWPOG
 EGPRRAGVSSFGVSGTNAHVILEQPGPDAAADAPDPTVDPGALAWVLARSNEAALRC
 QAAALLSVAGSDALCAXDIGHSLVTGRSSFAHRAVVMGQDRDALVRALSALAGEAD
 AGLAEGASAGRTAFYLFQSGSGRLGMWELYARYFPADAPDACAALDEHLERPLR
 DWGCEALINQATAYAQGLQFAIEVALYRLAESGMGRPDPFVAGHISGEVAHAHVSGV
 FSLPDACALVAQGRMLQOLPSGGAAMAI RATEDEVDPLHAEVSLAAVAGVSSVVIS
 GAEDVLAIAAHPAGEGKTRVLRVSHA FSPLEMPLEEFRAVTVRLSFCGTPTTPV
 SNLTGRLEPOLAHADTRVHRVREAVRPADGIQALRAEGVTRFLELPGDGVLSMAR
 ESASDDAVLAPLHADRPEETALLGALAQVLVYRGARVDMTVFPAGSGARMDLPTIYAF
 QHVRFPWSGVAGPDGVLRASGSLGASGLLGAALVELAGSGGLFTTGRLSVSESHPWLA
 HIFVLSVPGTALVRLVRAADGACOLLBELTLAALPLVPASGAQVQVAVGEEDP
 EAGRSPVHAREGSGPMTLHASGAVTSGAEVPFDATWPPKGAEPVDVADQVOLA
 DAGLYTPGAFHCLQALGVDAEAKLPSTDGDAYGLHPALFDAAHLASALGSAE
 AGGVFPWAGVSLHATGASHLKVRIREAGGALSVALADTSGAPVASVSLVIRPLSAG
 VQQAADRLDGLDFAVPSVAGTSGSQAQYVAGNAPFLDALA THRAAGLUPVSLA
 WNLHAEGLDGLDFAVPSVAGTSGSQAQYVAGNAPFLDALA THRAAGLUPVSLA
 WPGMSODGCTMTLSDAQVRIARQCMPLTVYEEGALFDALGSAESPMALPVRLD
 ALRAQCEQPTPLLRGLITRTRRSQAASAGI AQRLAGISTAEAREALLDVRAQATY
 LHACQETIPADRAFDQLGLDLSLTAIELNLLGKATGLRLPATTVDFYPTVDVRAATLH
 LDELFGAETGATETPLVPVGLPSLADDPVTVGMSCRPFGVAGSPEDLWRLVADGV
 AVSAPPTDRGHEIDOTYDPERGCAITRSGSFLHDAEEDPFEFGMSPREALTITDAQ
 RLLLETTWEALERAGMDPATLRSGSPVAGVVMHYDITLLSGRFPFGVQSGSGAGV
 ASGRSYTTFGFGPAVTVTACSSSLVLAHQAQSLRGECESLALAGVTVVMSTPLTF
 VEFSGRGLSADGCKKCAFDAAADQVWGAEGAGILVLERLSDARRNHRITLAVRGSA
 TQDQASRLTAPNGPAQORVIRQALASGLSADVDAVEAHGTGVLTPGDIYBAQALLA
 NYGQRPEDRPLLGLSVKSNIGHQAASGVAGYIKMVLAMRHGVLPRTLHVDEPSHH
 DWSAGAVELLTSEAEWPOGEPRRAGVSSFGISGTVNAHVILEQPEVPAETESSITPDT
 APDAEADAASGTPVALLSGRSASALRAQAARLLSRUDGDPGPRITDVAYSQLRG
 SAFPRAEVIILANPADRLHSLTALAEHGTTEAPVAVQAQDRASRGLATFVQSGSLOGL
 MGRELYGRYPFAEALDACAALDHLDRPLRDITWGEDAELLNTGYAQTLGFAIEV
 ALFRLYLSMGVPRDHLGHSIGSIAAARVAGVLSLPDACAALVAARGLMOOLPSGGAM
 MAIRATEDEVLPHLAEGVSLAAVAGVSSVVSAGDEVLALAAHPEEERKTKTRLRS
 HAFSPGDMALRFAVDADCHTYAAAPRIPVSVNTGRPATATAHPECCAEYVGHVREA
 VHPDVGMLRQGVTVLFLBLGDDXSALAAESAADSVLAPLVRKNRPEAPALLTA
 LARLHQGTTPVMSAAFACTGARVWDLPITYAFQHERFMPWSSGAARAGVRSAGLSAG
 HPLLGAAVRLGSGRGLLTGRLSLSHPMLADHLVGLSVLPCTALMEILVLRADAEVD
 CAAPDELTLAA PLVLPASGAAITQVVMWYCEPDEAGRRPVSVHAREGEGPWLHDAQGL
 APAATVETLPAALIMPOGAEHLDAAGCYERFADAGPAGYQFVQGLRAAKMLGDIYAE
 VALPBGTDGNAVGLHPALFDAAHLAALLGEGTDEAAVFPSSNGVTLHATGASRVVR
 IHPTEGTSIALDITVDTAGVPSVRLTARPTI TAGQTLGTDRLSFLPQDVMWTLHLITDER
 ANSLALLGHTDEGILDITLSQPHADLDLDAATGVHDTVLPAITPTFAGTVESVHAAT
 GALATIRLWLAADRFAASRLVFTVRGASCTDLGASVMGLSVALLHEHPGRFGLVDV
 DVYDQAEVLPRLALASDRPQVLVAREGLAARVRAQSSDTVTWDPGCTVLITGCTG
 GIGRSVARNHLSVEHGVRSLLLVRERGPAAEGVDALVALECAQVTVREACVTDVA
 VADLVARHISAVVHTGATDGVDSVESLTPRLSAVLIRPKVDAAWNLHBEATRGDLDA
 FVVFSGVAGTSGSQAQYVAGNAPFLDALAYTHRRVAGLPAVSLAWGPGWSDGMDTGL
 SDADVYIARQCMPLTVYEEGALFDALGSAESPMALPVRLDLAALRAQCEQPTPLLR
 LTRTPQRRITAAATEGDTAAATFAGRLTGLSAAEGREWLVGAVRSQAQVGLHAEATFI
 DODRAPLDIGFDSLTAVELNRLNGAVTGIRLPAITLLFDYPTPAELVAHLHARTAPEPT
 VQPEALGILERWEKSPFGDOLLTEENHEQIAGLEVLIRAKMDALRDTAAAHAGHDSFS
 DSDPFEASASDEDFVOLLDEGLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t
 ORIGIN

Query Match 45.6%; Score 112.2; DB 6; Length 20394;
 Best Local Similarity 56.1%; Pred. No. 4.1e-14;
 Matches 162; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

1 CTGGCCCATCAGACGTGGCGCCACCAACGGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60

EVVTEACDVTDAVAVADLVARHRSIAVHTAGVLDDGVVSELTPELRLAVLVRPKVDAA
WNLHTEATGGLDLDAPFVSSVAGTQAGQANYAGNAFLDALAYHRRPAGLPAVSLA
WPMQDQGTGLSDQVRIARQMPPLTVEGLALFADALGSAEPMLFVLRLDLA
ALRACQEPQLLGLIRTRRSRGAAGAGLQATGLSTAREALREALLVVRQIATV
LHAGPETIAPDPAQDLGLSLTAELRNLLGKATGLRPAITVFDVPTVDAALAH
LDELFGATGTATETPLPVGLPSLADPPVIVGMSCRPFGVSAPELRLVADGV
AVSAPPTDWEIIDDYPEREGAIATRSGLFHDAAEFDEPFEGMSPREALTDDAQ
RLLLETTWEALRAMPDPTLRGTGTGAGVMDHYSTLLSGREFEGYQGSAGSV
ASGRVYTFEGEPVATVTCSSVALHAAQSLRSGECSLALAGVTVMSPTPLTF
VEFRQGLSDAGCKAFADAGVGAEGAGILVRLSDARRNGHRLIATVRSVAV
NODGASGLTAPNGAQOQVIRQALASAGLSAADVAVEAGTGTCTGDPTEAQAALLA
TYGQRPEDRLLGYSKNSIIGHAAGSVAGVINKWLMARHGVLPRLTHVDEPSSHV
DWSAGELTSEAEPOGEGRRAGVSSFGISGTNAHVILELQEPVAAETESITPDT
APDAEEDAGSTPPALLSGRSASALRAQAALLSRLDGDPGPRITDVAYSATGR
SAFPHRAVILAAARADLHLSLAEGHTEAPVAVOARSGKLAFLFSQGSORLG
MGRELYGRYPAPAEALDAVCAALDAHLDRPDLVWGEDAEELNARTGVAOTGLFAIEV
ALFRLLESWGVPPDHLGHSIETAAHAGVLSLPDCAALVAARGRLMOQLPGSGAM
MAIRATEDEVPLHLAGVSLAAGVNSVVSAGAEDEVILAAAHFEEEGKRTILRVS
HAFHSPLMEPLADFRVADMGTYAAPRIPVSVNTGRPATAEELCAEYVWGHVREA
VRFDGVLREEQGVTTFLGLPGDGLSALAAESAADSVLAPVLRKNRPPAPALLTA
LARLHAQGTVPVWSAAAGTARWDLPTVAFQHERFWPSCGAARAGDVRSAGLSAG
HPLLGAAVELAGSGRLITGRLSSHPWLAHVGLSVLPVGTALMELVLAARADEVD
CAAVDELTLAAPLVPASGAALQVQVWGEDEAGRPVSHAREGEGPWLHADGAL
APAAETVPDITAIWPPQAEHLDACTERFADAGFAYGVPVQGLRAWKLGEIDYAE
VALPEGTGNAIGHLPALFADLAALHLLGGEGTDEAAVPPFSWNGVTLHATGASRVVR
IRTPEGGTISIALVDTAGAPVASVRSLTARPTAGLOQTGDRDLSLQVDMTTLHLTDER
ANSALLKGTDEGLDLSLQPHADLDLDAATGVHDTVLAPLPRTAGTGVESVHAATT
GALALIRSWLADDFRASAASRLVFTVRGAVSGTDLGASVWGLVRSALLHEGRCFLGVDV
DVDOAEVPLVPLRASDEPOVLVRGGEVLAAVLRAQSSSTVTDWSPSGVITGTTGVA
GLGRSVAKHVSHEGRLVLSRRGPAGBAGVDALVAELAEACCTGATVCACTVCTG
VADIVARRHSIAVHTAGVLDGVVSELTPELRLAVLVRPKVDAAWNLHTEATGGLD
FVFSVAGTQAGQANYAGNAFLDALAYHRRPAGLPAVSLAAGVMDHYSTLLSGREF
SDAQVIRARQMPPLTVEGLALFADALGSAEPMLFVLRLDLAALGSAEPMLFVLRL
LIRTPGRTAAATGDTAAAFAGRLTCLSAEGREVVLGAVRSQIAGVGLHARATEI
DQDRAFLDLGSDLSITAVELNRNLGAVTGIRLPATLLFDYPTPAELVAHLHARATEI
VGPBALLGELLERMEKSPFGLDITTEEMHEQIAGRLVLRKAWDALRDTAAAGHGDSPS
DEDFPSSASDDEVDLDDNLGLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t

ORIGIN

Query Match 45.6%; Score 112.2; DB 1; Length 20394;
Best Local Similarity 66.1%; Pred. No. 4.1e-14;
Matches 162; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 1 CTGCGCCATCAGACGTCCGCGCACCAACCGCGCGCATTCGTCGAGCTCGTACGAGACCAT 60
Db 14677 CTTCGCGGCTGTCCACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14736

Qy 61 GTGCGGCGAGTGTCCGCGCACCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db 14737 ATCGCGACGCTCTCGGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14796

Qy 121 GCATCTCGGCTCGATTACTACGCGCGCTCGAGTTCGGAACCTGCTGATCAAGGCAACA 180
Db 14797 GACCTCGGCTCGATCTCCCTGACGCGCGATCGAATCCGCTGCTCGGCAAGCGCAC 14856

Qy 181 GGACTCCGCTCTCTGCTCGCTGCTCTTTCGACACCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 14857 GGGCTCGGCTCCCGCGCAACACCGCTGTTTCGACTACCGCGCGGTGGATCGCCCGCG 14916

Qy 241 CACCT 245
Db 14917 CACCT 14921

RESULT 8
AX067996
LOCUS
DEFINITION
Sequence 3 from Patent WO007722.
ACCESSION
AX067996
VERSION
AX067996.1
KEYWORDS
GI:12329806
Streptomyces natalensis.

ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
AUTHORS Martin,J.F., Aparicio,J.F. and Colina,A.J.
TITLE Genes encoding enzymes in the biosynthesis of pimaricin and the application thereof
JOURNAL Patent: WO 007722-A 31-DEC-2000;
DSM N.V. (NL)
FEATURES Location/Qualifiers
source 1..20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
CDS 1..20394
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC24642.1"
/db_xref="GI:112329807"
/translation="MSNEEKRLKRAIADLHETROOLDTEAKQREPLAIVSMACR
PGCVRSPELWELLRDGVDAVSSFPNRCWDLDAVHSDPAHOGTSYARSGGFLHDA
GEDPGFEGISPREALAMDQOQLLLLETAHEAVERAGIDPESLGSRTGVFVTHGCH
YDAERRRDEVEGHLTHNHSIAGRSISVLEGGFALTVDACSSSLVALHLMH
ALRDECMALVGATVMSPTQMPFVSQRGLAPGCRKPPFAAADGTGSEGVGLL
LVERLSDAVRNGYPLAVLKGSVNDQSGNGLTAPNGPQQRVIRQALTGAGLAASD
IDAVEAGTGTGLDPEVAHALLATYQOARADRPCGLGSKNSIIGHTAAAGLVAM
KWLMAHGHGLPRTLHLDEPTGHVDMSEGNARLLAEPEWPSAGRPRRAVVSFGISG
TNAHVILEQAPAEAPAPAPARPGALPWLISARTAGRLAQDRGLRHRLDRADLE
FAVAHALADUTRILMEHRAVVVAGDREELRGLDAAAGRTANGLVAVKAAASFL
FAQCGSPQMGRLHAAHFEVTAFAVCAELDPHLDRLDRIDTVFAEESGAEAALD
QTYTQALPAGMAVVAQVEDIRERLACHEDHVAALAAAGPDSVTISGDEAVTET
RGTLMQALPAGMAVVAQVEDIRERLACHEDHVAALAAAGPDSVTISGDEAVTET
AAHWEAQRTKRLVSHAPHSPHMDLDEFRVARGLTTHAPRI PVVSTVTCALAT
EDELSPQYWRQVRETRFCAAVRLAEGVTTFFIGTGVGLTPWQDCLTTLEP
VLVPLRTGTPETVALTEGATVAFVHGVPVDSAFPAGTSGRADLPTAFQRQWYL
DRADHDEGEAAAAGEAGFMAAVEREDLOELSAVLAIDGSEADSLGSFLTSSWR
QRTQAAADRFSYRTHWAPRTAGGTATGHLVLPVPEGGTDPTWARTLALNDQGL
HTDVPADHEPDWGRHVPDGVCLLALDERPSTRSCPPYRGLAATTAARAEPEGA
GIOAPLWCVTGAVAVDRHEALKSPLOAOTWGLGRVAALESPOSWGLLIDLPNDLGR
AVSALLSTAGEEDQAVRPAVPAARLERITFGGTGDRWSTHGTVLTGTTGALGA
HLAHADAGAHLVLTGRRGPOAPGAPELAALTDGRVKVTLAACDAARDALAAVL
ADTLPPLPLTGTVVHAAGLDDGVLDALTPTFERFETVRPKAAQNLHLETDLDLHF
VLFSSVTVLGNAGQANYAANAAYLDALAEHLAQGLPATSVSNGPQQAAMHSDAA
DMRSRGLLLPMAAARRRPPAPALAQGTQVTVADIDMSAVAPALTA VPSPLGDLPE
ARRALPAEGPRESPRLDRICGALPAEQEKAFLTWREEARVILGHSPDPTVDAQR
AFREQFDSLMAVDLNRSLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
VQASTAALDEPVAIIIGMACRFPGVHSPEALWRLLAEGGDAITPMPADRMDLRLHV
PDPDHQTSYARGGFLDGAADFADFGISPREALAMDPOQLLLELTWELTEQAGID
PESLRGSSGTGFAGTNTQDGTALDAAQDEAGGRLTGNAMSVVSGRVSTFGPEGPA
LTVDTACSSSLVALHMAAALROGCSLA VAGGVYVMA TPSSFVEPARQGLAPDGR
KPAARADGTGSEGVGLLVERLSDAVRNGHQLVAVRQSAVNDQGSALNSPSP
SQQRVTRQALANARVAASEVDVAEAGTGTTLGDTIEAQLATYIGERPLLGA VAS
NLGHTQAAAGVAGVGMKWLAMRHMLPRTLHVDEPTGHDWMTAGAVELLTHEHTDWP
GHPRRAVSAFGISGTNAHVLELPAEQPLVEQPSAAEPDAPATAPDPTASDGT
PLLSAKPSALRAQAARLHSLHRLPALRLTDAAVTLMTHTRAFAHRAVPAVFAE
ALRLTALAAGEADPAVDTGTRDAVLFCGQSQRICMGRELSGRYPVFAEADPT
VCAALDEHLDRPLRDVVRGEDELNRTVYAQGLFAIEVALFRVLSWGRVRYVAG
HSVGEIAAAHVAGVFSLADACALVAARGLMQALPAGGAMAAIRATEDEVLHADS
TAAAGPSSVSVVSGAEHVALSTAHPGEGARKTTLRSHAFSLPMDPLMDLFRVAV
AELGTYCEPLAVSVNTQDLRTPEYVWTVHRAARFADGIRALGACGVTF
LELPGDGLSALARESADDAVCTPVLKDRSEAA TLAAHLHLVHGTEIDMTAFILA
GRDAHVDLPTVAFQHORFWPTDHTRTGDLGAVLEATGHPILLSAVELPDGEGLLF
TTFLSLQTHPMLAGHVMSVLLPGTFAELARAADAEVGCDDVDELTLA PLVLP
GGVQLRVGPADASGRRTILAKSRAEGDRLPWQVQHTAGVLAEGSTPPGVDFFHTE
SWPDAAPAVELSLYDPAAGHFDYGFPHQGLRTAWRRGDEFAEVALPAEAGEAS
AYGLHPALLDAAHLGVVAFNGVDVRGVVFSWESVALHATGASAVRIRVPHSGDTSV
VADTTEGPVSGITLVRVASDQAGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LQGLSACAPDGYDPLASLARSGRCAVLPVPEAGAEVVDVVGATAGARALD
ARSLADDRFAASRLVFTVRGAVSGADLAGAAWGLVRSALSHPGRFLVLDLDDAE
LALVPRVSDSEFPQLLRGVEVLAAARLQAQSSHAVITWDFSGTVLTGTTGGGGRVMA
RHLVVEHGVNRLLLSRRGPAAEABELVTELRHSGAEVAVEACDVTGAAADVAV
HRSIAVHTAGVLDGVVSELTPELRLAVLVRKAWDALRDTAAAGHGDSPS
AGTIGSPGQANYAAGNAFLDALAYHRRPAGLPAVSLAAGVMDHYSTLLSGREFEGYQGSAGSV

GTRADAPAGPSWRTDITVLVTGCTGALGAHLAQWLAAGAAHVLTLTSRRGPADGGA
 AELTLRLRETCETVTAACDVTDRDLAELLSGLPADRLTGVVHAAGVLDGVLDSL
 TPDRAFAVAGPKVTGAHWHLHETJDLDLASAFLVPSFAGTGVLAGQAANAALVLA
 LAVRAOLGLIPATAVAGWSGAGMAADTEARQLARTGLVPLPAALAAAGVMD
 DGETAVTVADVMDRPAAGFGPGRPSLLTGVPEVRIERETGRAGAPAEITDRGLGL

Query Match	46.11%	Score 113.4	DB 1	Length 104326
Best Local Similarity	67.71%	Pred. No. 1.6e-14		
Matches 159	Conservative 0	Mismatches 76	Indels 0	Gaps 0
Qy	1	CTGGCCCATCAGAGCTCGCGCGACCAACGGCGCGCATGCTCGAGCTGCTACGAGCCAT 60		
Db	83773	CTCGCGGAGCTGATCGCGCTCGAGCGGAGGAGACCTCGTGGCCCTCGTACGCGCGG 83714		
Qy	61	GTGGCGGAGTGTCTCGCGACCGCGACCCGAAAGCATCGCGCGCGACCTCGTTCGCT 120		
Db	83713	ATCGCGCGCTGCTCGGCGACCGCTCTCGACCGGATCGAGCGGACCGCGCTTCGCG 83654		
Qy	121	GCATCTCGCTTCGATTCTACTCACCGCGCTCGAGTTCGGAACCTGCTGATCAAGCGCA 180		
Db	83653	GACCTCGGCTTACCTCACTGACCGCGTGGAGCTGCGCAACCGGCTGAACCGCGGACC 83594		
Qy	181	CGACTCGCGCTTCTGCTCGTCTCGTCTCGACCGACCGACCGCCCTCGCAACTCG 235		
Db	83593	GGACTCGGACTCGCGCTGAGCTGCTTTCGACTATCCGACCGCTGGCGCAACTCG 83539		

RESULT 7
 SNAL13222 20394 bp DNA linear BCT 24-APR-1999
 Streptomyces natalensis pims1 gene.
 DEFINITION
 ACCESSION AJ132222
 VERSION AJ132222.1 GI:4678702
 KEYWORDS pimaricin biosynthesis; pims1 gene; polyketide synthase.
 SOURCE Streptomyces natalensis.
 ORGANISM Streptomyces natalensis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Streptomycetales; Streptomycineae; Streptomyces.
 REFERENCE 1 (bases 1 to 20394)
 Aparicio, J.F., Colina, A.J., Ceballos, E. and Martin, J.F.
 The biosynthetic gene cluster for the 26-membered ring polyene
 macrolide pimaricin. A new polyketide synthase organization encoded
 by two subclusters separated by functionalization genes
 J. Biol. Chem. 274 (15), 10133-10139 (1999)

JOURNAL MEDLINE 99214571
 PUBMED 10187796
 REFERENCE 2 (bases 1 to 20394)
 Aparicio, J.F.
 Direct Submission
 Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.
 Real, 1-Leon, 24006, SPAIN
 FEATURES
 Location/Qualifiers
 1..20394
 /organism="Streptomyces natalensis"
 /db_xref="taxon:68242"
 gene 1..20394
 /gene="pims1"
 CDS 1..20394
 /gene="pims1"
 /function="pimaricin biosynthesis"
 /codon_start=1
 /transl_table=11
 /product="polyketide synthase"
 /protein_id="CA041041.1"
 /db_xref="GI:4678703"
 /translation="MSNEEKUREYLYKRAIADLHETROQLDETEAKQREPLAIVSMACR
 PFGVRSPEELWELLRDGVSSPPRRNRWDLDALYHSDPAHQTSYAREGGFLHDA
 GFDFPGFICISPREALAMPQORLLLETAWEAVERAGIDPESLAGSRPFVFGTHGG
 YDAEGRRADEVGGHLLTGNHSIAGGRISYVLGEGPALTVDTCSSSLVALHLMH
 ALRDECAMALVGTAVMTQMFVPEPSRQRLAPDRCPKFAAADGTGSGVGLL
 LVERDSARNGYVTLVKGSAVNDGASNGLTAPNGPSQVROIRQTGAGLAASD
 IDAVEAHTGTGLDGPVEAHLATYVQQRADRPCGLGSKSNIGHTQNAAGTAGVM
 :KMLVLMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPPEPWPSPAGRPRAAASVSGISG

THAHLVEQAPAEAPPEPAARPGALPMWLSARTAGLRAQADRLGRHLRDLADLE
 PAAVHALADTRTLMEHRAVWVAGDREELRGDLDAAGRTANGLVGVAVKAASAPL
 FAGQSGPCMGRELHAHPVPATADPAVCAELDPLHLPRLDRIVFAEESAEAAALLD
 OTYTAOALFALETALFRLVESMGVAPRVAGHSIGELTAHVSGVTLTHDAARLVAA
 RGTMLAALPAGGMVAVQATEDIRERLAGHEDHVALAANGPDSVTVISDEQAVTEI
 AAHWEAGRTTKRLRVSHAFSPHMDMLDFRRVARGLTFLHAPRIPVSVTVTGALAT
 EDELSPDYVWVQVRETFCFAAVRTLEAGVTFVFEIGTGGVLTMPVOCQLTLBLEP
 VLPLRLTGRPETVALTEGVATFVHGVPVDSAFPAGFAGTSRADLPTTAFQWQWML
 DPADHDEGEAAAEAGEAFMAAVEREDQLSEAVLADIGSEADSLGSLPTLSSWRR
 QRTQAAADRFSVRTHWAPRTASGPTATGHWLVLPPEGCTDPMPTARLLDALNDOGL
 HTVRELPAHDEPDANGRPVDCVLLCLALDERPTRSCPPYRGLAATNNAARPEGA
 GIQAPLWCVTGAVAVDRHEALSKPQAOQWGLGRVALESPOSWGGLDLPNLDGR
 AVSALLSTLAGEEDQVAPAGVPARRLERITPGDGTGRWSTHGTVLVTGQWGLDGL
 HLAMLAGAHEHLVTGRRGPQAPALAAATDRGVKTLAAACDAADRDLAALV
 ADIPLHLPTGVVHAAGVLDGVDLALTPERPETVLRPKARAQVLEHTQDLDLHIF
 VPSSTVGVGNAGQANYAANVLDALAEHRLAQLPATSVSMGPGQAAAMHSDAA
 DNRSGLLPMAAARPRRPAPALAOQMTQVTVADIKSAVAPALTAVRPSPLIGDLPE
 ARRALPAGCPRRERSPLKDRIGALPPAQEKALFTWVREEARVLCGHSVPDVAQR
 AFREQGFDSLMVADLRNLSAATGLRLPATLLPHPTPLAAACALSLRSEVLGAAGPATV
 VQATTAALDEPVAIIGMACRFPQGVHSPEALMRLLAEGGDITMPADRGMDLRLYH
 PDPHOCSTVARGGFLDGAADFPGI8PREALAMPQORLLLETWELVLEAGID
 PESLGSSTGVFAGTNTQDYGTALDAADQABAGHRLTGNAMS VVSGRVS VTFGEGPA
 LTVDTAGSSSLVALHMAAOLROGECSLAVAGGVYMATPSSFEFAROGELAPDCEC
 KPFAAADGTGWSGGLLVRLSDARRNGHVLA VVRGSAVNOGASNGLSAPSGP
 SQQRVIRQALANARVAASEVDAVEAHTGTTLGDPLEAQALLATYGOERPLLGAVS
 NLGHTQAAAGVAGVMKVMAMRHGMLPRTLHVDEPTGHVDWTAGAVELTHTDWPET
 GHPRAAVSAGISGTNAHVLELPAEQPLVBPQSAEADAPATAPDRPTTASDGT
 PLLLSAKESALRAQAARHSLERDPAALRLTDAAYLTMTHTAFARAARVAADHEA
 ALLALTAAGEADPAVDGTGTAHTGRDVLPSQGSORLGMGRELSGRVFAEAPDT
 VCAALDEHLDRLPDRVVRGEDELLNRTVTAQAGLFAIEVALFRLVESWGRVHYVAG
 HSYGEITAAAHVAGVFSLADACALAAAGRLMOALPAGMAAIAATDEVLPHLADSV
 SIAAVNGSPSVVSGAEHVLSTAAHFEAGARKTRLRVSHAFSLMDPLMDLFRV
 AEGLTGPELAVSVNTGOLATPDQLTPEYVWTVHRAVRPADGIRALGAGGVTRF
 LEIGPDGVLALARESADDAVCTPVLKRDSEAAATLLAALTHLVHGTGIDMTAPLA
 GRDAHVDLPTYAFQHQRPMTDPDTRTCDLGVLEATCHPLLSAVALPDEGGLP
 TTLSLQTHPLAGHVMSGLPPTAFAPAEALRAADVCCDRDDELTLAAPLVLPBH
 GGVLQRLVGPADASGRRTLTARSRAEGDGRPMVQHTAGVLAEGESTPFGPQDHFTE
 SWFPADAPVLESLGYPDFAAHGFDPHFQGLRTAWRDEGVFAEVALPAEAGEAS
 AYLHPALLDALHVAHFNFGVDRGVVPSFWSVALHTAGASAVIRVVRHSIGTSVD
 VADTGPVSAICTVLRAVSADLQAGGADPAVRDALFRVQMPNPLRPAAGAVTAT
 LGSLAGPDGYDPLASLARGVAGNVLVPVEAGGEVVDVVGATATARALDL
 ARSLADDPRAARSLVFTVKGAVGADLAGAAVGLVRSALSBEHPGRFLVLDLDDAE
 LALVPRVLASDEPOLLVGGEVLAARLAQAQASHAVTMVDPSTVLTGTTGGGLGVNA
 RLHVGEHVRNLLLVSRGPPAAGEABELVTELHSGAEVAEACDVTDAAANADLVAR
 HRTSVVHTAGVLDGVESLTPELSAVLRPKVDAANLHEATRDLDLDAVFFVSS
 AGTISPGQANYAGNAGLDALAHRRRAAGLPAASLAWGPMWSDGGTGTLDVDS
 SPGRHARTPTGRLPDAALAAGDAHLVPVDFWASLRAQGEVPLLRGLLIRARR
 SAVGSSAAAGLVSRGRTVERVELLDLRAQIAVLGHANPETIESTRVPQDGLG
 FDSLTAVELNRNLNATGLRLSATVDFYADALVDLDELDFGAQSEALPAPVPS
 PAGAADPVVIVGMSCRYPGCGSPEDLMRLVSEGVDAVSDFPTDRGWDESLSPDP
 EALGTSVTRSGGLHEAABPDPDFGMSPREALATAOQRLLETWELTWEAIBRTGIDPA
 SLRGSRTGVAGVMTYDGLLVGDQPEGVRSNGSRAASIASGRVSYTFGEPGPAVTD
 TACSSILVALHWAQSLRSCESLAVAGGVTVMTPTTFVPEPSRQGLSDAGRCAKAFA
 DAADGVWGEGVGLMLERLSDARRNGHVLA VVRGSAVNOGASNGLSAPNGAQQ
 VIRQALASGLAADVDVAHAGHTTGLDPIEQAALLATYGOERDEPRLLSGVS
 NIGHAQAAGVAGVIMVLAHGHVLPRTLHVDEPSPSHVDSAGAVLLTSEAEWPG
 EGPRAGVSSGVSNTVILBOPGDDAADAPATVTDGPALAWLSARNEAALRC
 OARLLSLVAGSDALCADIHLSILVTCRSSFAHRAVVMGODRALVRLALGAEAD
 AGLAEGSAGRTAFLPSGCGSORLGMGVELYARYVPVADAPDAVCAALDEHLEPLR
 DVVWGEDAELLNTAYAGLFAIEVALYRLASSWGRPDPVAGHSIGEVAAHVSQV
 FSLPDACALVAARGLMQQLPSGGMMAIATATEDVLEPHLAEGSVLSAANGSPSWTS
 GAEDAVLIAAHPAGCKRTTLRVSHAFSPHMDMLBEEFRAVTRTFSFGTPTTPV
 SNLTGLRLAEPQLAHADYVVRHRAVRPADGIOALRAEVTFRFLELPGDGLVSAMAR
 ESASDDAVALAPVLRDREPEETALLGALAOYVRGAHVMTVPAGSARVADLPTYAF
 QHREFPSPGVARPDVRSAGLSAGSLPGLCAAVELAGSGLLFTGLRSLVSSHPLAD
 HVLGVSVPGLTALVELLADAEAGCDLLELTLLAAPLVLPASGAAVQVQVAVGEPD
 EAGRPVSVHAREGEGETWTLTASGAVTSYSGAEVFPDFTWMPKGAEPVDVADCVLA
 DAGLTYPAGHGLQAQWAGVYAEAKLPSTDGDAYGLHPALFDAALHACSDYGLAE
 AGVPSPWAGVGLHATGASHLRVRIEAGGALSVAIADTSGAPVASVSLVIRLSAG
 VQQAADRDALPKADHVPVPLTDERVEPTGCEPEPLRTYADLDSLEGAAVPGTVLVA
 PSQAADRSVSHAATVWALEMQRNLADDFATSRSLVFTVTRGAAFGADLAAAARGLV
 RQAQSENPFGLVDMDMDGTTPVPAQALATDEPVLVGRGVEVLAAHVAARQSSHTVT
 WDPSTGTLLITCGTGGLGRSVARHLVSEHGVRSRLLLVSRGPPAAEGAGELVAELRGS

entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase, PF00698 Acyl transf, Acyl transferase domain and PF00550 pp-binding, Phosphopantetheine attachment site and matches to prosite entries PS00606 Beta-ketoacyl synthases active site, PS00340 Growth factor and cytokines receptors family signature 2 and PS00012 Phosphopantetheine attachment site"

/codon_start=1
/transl_table=11
/product="putative type I polyketide synthase"
/protein_id="CAC22144.1"
/db_xref="GI:12274801"
/db_xref="SPTREMBL:Q9EX54"
/translation="MSNEOKMVEYLKWVTTDLQKARORLAEAGNDEPVAIVGMACR
FPGVASPDLLWRLVMEADAGITDMPDTRGDWDIDAIYDPERGPKGKYTRKGFPLDGA
AQDFPGFISPRALANDPQORVLETTANETFEQAGIDPATLRSQGVGVPAGVVEGS
YLGLEPELEGYLTWSKLVSASGIATYTFEGEPVAVSDTACSSVALHVALQSV
RKGSTWALAGSGTVSGNFGVDFSRQNGLAEDGRIKFAAADGTAWSEGVGLLV
EKLSDARNRHRLVAVGSAVNODGASNGLTAPNPSQERVIHQALAGARLNASD
VVEAHGTGRLGDPTEAQLLASVGRORPADRPLYGLSKSNIHGTVAAGAGGVIKM
IMNRHGLPKTAVNDPPTMVDMEAGAVELLTEARPKPTGAPRRAGVSAGVGTN
AHVLEPAPSEPEPADTAVLPVWALSARTAPALRQARLLSHVEERPAEDP
LDVYSALAESLEHRAVVVSDRELLTGLRALAGRTTPTGAVDTRTGKTAFL
FTGCAQTGMDLYRTFPAYAHAFDTITAQDPLHLDQPLHHTITTCGHLHHTGNTQ
PALPATEVALYRLLETWGTITPDYLAGHSIGELTAHISGLITLQDATTCLVARALMQ
NLPTHGMIALQATEOEITLPHLTGHEHLLTAAINSPTSLVSGNOTAAOHIAIOLQA
OGRKTLTVAQSHFHSMDMLHDFHTATQLYTHEPTIPIVSLTGLNLAHNDLT
FTYWDQUNTVRTYQAIHTAGVITYTTEIGPDATLTPLOTQTLTALPLLRN
QPEHTLLTALSRHNGVPIWKNLFTTTPAKHTPLPTAFQHERFWETSVGAADA
SDLGESAGHPMLGAATVAGADQVLTSLSLRTHPWLADHVLVDOTLVPATLVM
AIRAGDELRLVNDLSVLPVLPKGQHVQVSGSPDASGLRSLISVSPDNADV
PWTNARGLSFRGPDAPFSLAOMPAGAEVPLDEAYDKLAVGVYVGVPRGLEHV
WRLGDEVYADVRLPDGRTGIGFGLHPALLEALASCLAGYSSPEGAWISDVK
DVRGLHADRRLVRLPADGALNVQLADRAGRPVAVSVACRAIAGVSNALAR
NRDLFGVDMAPFPVAPSGATERWVSLGADFGARRPVDVASLARAAGSVSDVALV
GTETGTDGRVADVDVHTREALVQVMEARLGLVLRVTRGVTAGDEVDVTL
GAAAVGLLRQAENSGRIVLVDVDDDEASARILPAVVASGEPQSVARVQGVPRPL
GRPEAPAGATGAWDRGTGLITGGTSGGLFARHLVVEHNGVRLNLLTSRRGRDAE
GAABEALGTALGARVEIAACDAADRALAALADIPAEVPLTGVLTAGVLDGLIG
SQTQRSLAVLRPKVDATNHLTRMDLSAPLVFSSIAAVGPGQSTVSAANHEL
DALAEHRGLAATSLAWGLWSQTSQMTGHLDEADLKIARSFRPVESQGGPALLD
IALTVRAGLVAITMDLTALREQVPLVLSKLARTPVRTAQAAYSESLESLAG
LGDAREAVLIGVQAAYVHGSIDATGIDQRSFTALGDFSLTSVELRNQLGTG
LRLPPLTFDHTPVAAEYLHGEVRAELSGASRPVDFAAEVLADVRPADEV
ITTAADPKELLTGSAGFLGAPLLRDLMTTARVCHLVRGADENAMERLKAKEY
RWDDIDPERSVVLGDIAEPFLGDESPALARTVDVVTNNGARVHLPYETLKA
SNVGTDEIRLALAHRTVPVYVSTVGVDFGVEPGLVPLRTPDTPGAEISPGYLK
SKWAEQVIEVARDRGLEPVSRYVDVIGSDRVNGACQTRDFVWLTGLKLIQARSVPK
TEGRFHLPLVDVYSAATIGISRPQCTVGRTFHLFNQSSLSQCVELLRSGLYELDEV
DWDRTQVVTGSDNALLPLDFAFEMWTSDDTGDFFPIDTAETVAALGEGIGIPVLR
ELFKYVAFVVEGHFPAG"

misc_feature complement(4732..4935)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 68.50, E-value 2.5e-17"
complement(4810..4857)
/gene="SC2C4.03c"
/note="PS00012 Phosphopantetheine attachment site" complement(5248..5268)
/gene="SC2C4.03c"
/note="PS00340 Growth factor and cytokines receptors family signature 2" complement(7342..8286)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00698 Acyl transf, Acyl transferase domain, score 248.50, E-value 9.5e-71" complement(8596..9867)
/gene="SC2C4.03c"

Query Match 50.2%; Score 123.4; DB 1; Length 24225;
Best Local Similarity 69.0%; Pred. No. 1.8e-16;
Matches 169; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 1 CTGGCCCATCAGACGTCGGCCGACCAACGCGCGCATTTGCTCGAGCTCGTACGAGACCAT 60
Db 15698 CTGGCTCGCTGTTCGGGGCCGAGCGGCGACTACGGATCTCTGCGCGCTCCAG 15639
Qy 61 GTGGCGGCAAGTGTCTGGGCAACGCGGACCCGAAAGCACCACCGCCAGTCTGTTCCGT 120
Db 15638 GTGGCGGCGTGTCTGGGCACTCTGACCCCGGCGCATCGAGTCCGCGCGGCTTCCAG 15579
Qy 121 GCATCGGCTTCATTCACACGCGCGCTCGAGTTCGGAACCTGCTGATCAAGGCAACA 180
Db 15578 GAGCTGGCTTCGACTCACTGACAGCGCGTCTGAACTTCGCAACCACTGACACCGCGAGC 15519
Qy 181 GGACTCGGCTTCTCTGCTCGCTGGTCTTCGACACCAACCGCCCTCCCGCGGCTCTCGCGCC 240
Db 15518 GGACTGGCTTCGCCACCACTCTGCTTCGACCACTCCCGCGGCTCTCGCGCC 15459
Qy 241 CACCT 245
Db 15458 CACCT. 15454

RESULT 6
AB070940/c
LOCUS AB070940 104326 bp DNA linear BCT 22-JAN-2002
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION AB070940
VERSION AB070940.1 GI:15823967
KEYWORDS Streptomyces avermitilis DNA.
SOURCE Streptomyces avermitilis
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Oonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
REFERENCE 2 (bases 1 to 104326)
Ikeda, H.
AUTHORS Direct Submission
TITLE Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1 Shirokane, Minato-Ku, Tokyo 108-8641, Japan
JOURNAL (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp. Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1 Shirokane, Minato-Ku, Tokyo 108-8641, Japan
AUTHORS (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp. Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)

FEATURES
source 1. 104326
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
gene complement(1..1134)
/gene="mcma"
CDS complement(1..1134)
/gene="mcma"
/note="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=11
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
/translation="MTGISTAADANARFRLVSGTFAVPVFDLPTRAGLSDAPAA
RLVGRGVADVSDIDMRVLFGLRDLRVVPSLAVDAPAPLLVLYQLVAEEHGAPO
RLAGAVRGDFVEFLFARPGFPLPSLRMANVFAVCQALPRWQCLSCVGHHLAQT
GAPALEVAFVAGGIEYLRVAVTAGLDVTRVAPRLTCLVGGTTGRENRAKLATRR
VWTRTANERSGSRTAPRLVTRPDGLATLACTEDIEAGATLVADRIERLGGTLTVL
EHGLAGQLGRPHGDRATPRPHPHQDSLDLQAROTERLAKLRMRVQPAVDEALLR
LRKAAQDDNVLFPMLAARATVGVCDALRSTWGTGFSAT"
complement(1131..1886)
/gene="ppta"
complement(1131..1886)
gene
CDS

QY 121 GCACCTCGCTTCGATCTACACGGCGTGCAGTTCGAAACCTGCTGATCAAGCAACA 180
 Db 6155 GCACCTCGCTTCGATCTACACGGCGTGCAGTTCGAAACCTGCTGATCAAGCAACA 6214
 QY 181 GGACTCCGCTTCCTGCTCGCTGCTTCGACACCGACCGCCCTGCCAACTCGCCGTA 240
 Db 6215 GGACTCCGCTTCCTGCTCGCTGCTTCGACACCGACCGCCCTGCCAACTCGCCGTA 6274
 QY 241 CACCTG 246
 Db 6275 CACCTG 6280

RESULT 5
 SC2C4/c
 LOCUS 24225 bp DNA linear BCT 12-MAY-2002
 DEFINITION Streptomyces coelicolor cosmid 2C4.
 ACCESSION AL512902 AL645882
 VERSION AL512902.2 GI:20520683
 KEYWORDS accA1; secreted FAD-binding protein; type I polyketide synthase.
 SOURCE Streptomyces coelicolor A3(2).
 ORGANISM Streptomyces coelicolor A3(2).
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J.,
 Kinashi, H. and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 Mol. Microbiol. 21 (1), 77-96 (1996)
 97000351
 8843436
 2 (bases 1 to 24225)
 Seeger, K.J. and Harris, D.
 Unpublished
 3 (bases 1 to 24225)
 Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
 Direct Submission
 Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK
 On May 9, 2002 this sequence version replaced gi:12274798.

NOTES:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
 details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nhn.hgo.jp/jun/cgi-bin/frameplot.pl>.
 CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence

overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2C4.
 FEATURES
 source
 1. 24225
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid 2C4"
 complement(2. .1477)
 /note="previously sequenced DNA fragment. EMBL:AF113603 Streptomyces coelicolor putative acyl-CoA carboxylase complex A subunit (accA1) gene, complete cds."
 2. .119
 /note="nominal overlap with Streptomyces coelicolor cosmid SC4H10"
 complement(374. .1471)
 /note="Pfam match to entry PF00289 CPSase_L chain, Carbamoyl-phosphate synthase (CPSase), score 585.30, E-value 3.8e-172"
 complement(602. .625)
 /note="PS00867 Carbamoyl-phosphate synthase subdomain signature 2"
 complement(1485. .1489)
 1774. .1778
 1786. .3438
 /gene="SC2C4.02"
 /note="SCO6272"
 1786. .3438
 /gene="SC2C4.02"
 len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222) Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores: opt: 1323 z-score: 1315.4 E(): 0: 45.3% identity in 470 aa overlap. Contains Pfam match to entry PF01565 FAD-binding_4, FAD binding domain. Also contains possible N-terminal region signal peptide sequence"
 /codon_start=1
 /transl_table=11
 /product="putative secreted FAD-binding protein"
 /protein_id="CAC22143.1"
 /db_xref="GI:12274800"
 /db_xref="SPTREMBL:Q9EX55"
 /translations="MTEVSRRLKMGAAVSGGALPALGAPPATAPAPAGPDLPGP AAAAGRRPAPVHVTRSDPRTYQGLISGNQWGNPDYIRVVSSAGQVRAQVEAV DGLKTAIVSGCHDEDFVANRDVVIDMAGDVSVDYDRERAFVGPARGLTGTVR TLKRVGVLPGGCTPFGAGHITGGYGCALSRSGLTVDLHYLVAVVHVVDARGRR KVVATREDDPNRELWHAHTGAGGNFVITRYLWLRSPDVGDDPFTLLPAAPSEVLL SDVSWDDLDLEASFTRLLNFTFWHARNAPSDSPGFLFSQKTHKAAGYFRMSTQ VDAAPADRLDDYLAALISEGTVTHVGDYRAPWLYAVTWSGFEASVPRKSK SAVRVEMPEQLRAVYRQITRDDYPCPYGMIAIVGGGKINEVAPGDTATAORDSIA KMVCSLWSDPADALHQRWIREAYEDIVYASTGTVPRPGVNDGCIYINADADLADPA LNRSGLFWHELYFGNYPRLQVRKATWDPNRFVSHRLGIRKP"
 1948. .2535
 /gene="SC2C4.02"
 /note="Pfam match to entry PF01565 FAD binding_4, FAD binding domain, score 6.50, E-value 1.1e-06"
 complement(3514. .9972)
 /gene="SC2C4.03c"
 /note="SCO6273"
 complement(3514. .9972)
 /gene="SC2C4.03c"
 len: 2152 aa; N-terminal domain almost identical to, TR:Q9J8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2) type I polyketide synthase (fragment), 720 aa; fasta scores: opt: 4356 z-score: 4494.6 E(): 0: 93.8% identity in 721 aa overlap and C-terminal domain similar to SW:HEMTANASP (EMBL:L22883) Anabaena sp. polyketide synthase HetM, 506 aa; blastp scores: Score=761 (267.9 bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identifiers= 184/491 (37%), Positives= 278/491 (56%). Also highly similar, in this same cosmid, to the N-terminal region of SC2C4.04c and to SC2C4.05c. Contains Pfam matches to

misc_feature
 gene
 CDS

EVSRFRWDAVARQDLTVATTLAVPSAGLDTVVVPALSAWHRRHQHDQARINTWTVQST
WKSELTTHTRQQAQHWLIAIPETIGTHHPHTNLTNLNHHGHTTCTLLNLTHTNTPQI
LHHTLHTRQQAQHWLIAIPETIGTHHPHTNLTNLNHHGHTTCTLLNLTHTNTPQI
PPTPLVATYNTATTHTRQQAQHWLIAIPETIGTHHPHTNLTNLNHHGHTTCTLLNLTHTNTPQI
LQHLTQTLLTPQHQTQLAIPHTCTHTRBLLPTTTLTTHQBPPTPHGHTTLTGCTGAL
ATHLHTQTLLTPQHQTQLAIPHTCTHTRBLLPTTTLTTHQBPPTPHGHTTLTGCTGAL
QLNLTIPQHPHTTTHHTAGLIDATNTLTNLNHHGHTTCTLLNLTHTNTPQI
TAFVLYSSAAATGAPQOANYANAYLDALAHRRHTHLLPATSIANGTWQNGGLS
DKARAYLDHGRFMRPELATAATQAIADTREPYYVIADIDMSKTEHTSQKSLADPOS
AREREPAPQRTPPAEULHLLAQTSDQRAAELLVRDHVAALRHADKQIAIDPOS
FRALGFDLSITAFVFRNLLIKATGLRPLVSLFDPHTPAKLAHLQNLRGRTAESAAPS
AAVTAASVETPIAIVMGARPGGVTSADDFWDLISSEQDAIGCGFTDTRGWDLLTDL
YDPDPHGCTYTRNGGLYDAGHDAEFGISPRFADLMDPOORLLETTAMETIETHA
GINPHTLTGTPVGTCTNGLQYALRVHNAAGSTDGFALTGTAGSVISGRISYTFGE
GPVAVDTATSCSSILYALACALRAGECSMALAGTVVMSFGAFVSRQGLFAD
GHCKAFSAAGDGTGWSGVMLVERLSDAHRNGHRLVAVRGSANVDQAGSLTAP
NGPSQORVIRQALANAGLSAGLDVDAVEHGTCTTLGPTIEAQAALLATYQDRAEGELP
WLGSVKSNVGHTOQAAGVAGVIKNMALRHGLLPRTLHVDEPSPHVDMAGAVQLLITE
TVPWPGEGEGRURRAGSVQSGVGNHVLIEEPADDPGGPAGSGDAGSDDEAAAA
SPGVPMVLVSAKQSPALLRAQAALHAHLTDHPGLDLADVGYTLAHARAVDFHRATLIA
ADRTDTFQALQALAAAGPHPAVIHSSAPCGCTGEAAGKTAFTICSGQGTQPGMAHGL
YHTHPFAAALLNDICTHLDPHLDHLLPLLTQNDNNDNEAALLQOQTRVAQPAFAPQ
VALHRLITDGVHITPHYVAGHSGEITAAHLAGILITDATTLLTORATLMTMPBGT
MTLTHPHSHHTHLLTAHENDLALAIANTPTSLVISGTPHTVQHIITLCOQQGIKTKT
LPTNHAHSHTHNTNLNQLHQTALTYHPHTPLTITANTPPDQLTTPHYTQOAKRT
VDYATTTTTLHQHGVTYYYIELGPDNTLTLTHLNPPTTTLTTPHHHPQTHLLT
NLAATTTTTLHQHGVTYYYIELGPDNTLTLTHLNPPTTTLTTPHHHPQTHLLT
TEHPLIGATLIEATDGCALLDAGLSLRSHHPWLAADHAVGGTVLLSGATFLELALHAGTY
VGCDRDEBLTILHAPLVPVPGDAGSVQVGVAAADGEGRRLVSVYARGSGSGGGGAGG
VWTCASGLVEAAANGVGVVDGLAVGPPRGAVADVGVDRDLRAGACVGLGVFSGL
RAYWDRDGLLAEVGLPEEANGVGLRHPALLDGVVQPLSLVPGGTGFGGAGFG
RAGVRVPAWGVGSLHRRAGTVGRVRSAGVGRGREAIVSVGVDEGAPVSVDRLEL
RPVMDQGLRAVSVSAGRRGSLYAVQMAVGVPVPCQQAWEHEDVCSGGSPVGVVV
LRCPDAGAGGGGGGGGVGVGVGVGVGVGLGLERFAGSRLVSVYTRGAVACPE
DGPVVDVGSVWGLVRSQAERHDPDRFVLLDLDTDTDTDTDTGAGAGWVDCGRVAWV
ACGPEVRLARGLRLLARGLRSGSDVPAQRSDTKRRSDVLRSGSDVPAQRSDV
VSGREVLPMVSGGSLVVTGTVGLGVAARHLAGVCDRDLVLRSGSDVPAQRSDV
AELALAGAEVRIVACDVGEBRVRVRLGVPAGCPLTVGVHAAAGVLDLDTASLTPR
LGTVFAAKVDAALLDELTRGMELSSAFVLSAAGILGSQAGQNYAANAALLDALAYR
RRAAGLPCVSLAWGLWEASGMTCHLAGTDHRRIRISGLHPMTSTPDALAALFDAALLD
RPTLLVPRHTPAPPLPQLDQLLTPARRTRRTTTTGGADQAQLHRLAGTQHEQD
TTLVLRADRTATVLTHTTPTTTPDRAFRDGDSLTAVELNRGLTGLRKLPTTL
AFDPHNPHTTHLHTLQLOPQDNNAVPALEALDKUESALSALDKTDSASERTVLRK
SLMLRWNAQPHPTAESADDEKFTSATSEABIEFKPIDNDLGLS"
12071..30790
/gene="avea2"
12071..30790
/gene="avea2"
/note="contains Ave module 3, module 4, module 5 and
module 6
multifunctional polyketide synthase"
/codon_start=1
/translation_table=11
/product="type 1 polyketide synthase AVES 2"
/protein_id="BA084475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLAWHLVKSQTDALIGESTDRGNLEOLYDPPDRSGTSYTRSGGLY
DAGDFDAAPFELSFPREALMDPQORLLETWTETFEQGGIDPRSMGRSGRTGVFVGINP
EDYTYGTHQPSNAVGLLGTGSAISAGRSISYFNGLEPAITDITACSSSLVALHL
ACQALRGECTMALAGASVNAFTVFETFSRQGLAADRCKFAASGADDTGCGSEGV
GMILVERLSDAHRNGHRLVAVRGSANVDQAGSLTAPNGRSQVYIRQALANAHLS
PADVDAVEHGTCTTLGDPTEAQAALVEAQQDRNGRPLWGLTKSLNISMMAAGV
GVISGMALRNGLLPRTLHVDEPSPHVDMAGAVQLLITETVPHGEGRLRAGVSSP
GVSGTNNAHVLIEEPADDPGGPAGSGDAGSDDEAAAAAGSGDDEAAAA
KSPQALRAQAALHAHLTDHPGLDLADVGYTLAHARAVDFHRATLIAADRTDLFQALQ
ALAAAGEHPAVIHSSAPCGCTGEAAGKTAFTICSGQGTQPGMAHGLYHTHPFAAAL
NDICTHLDPHLDHLLPLLTQNDNNDNEAALLQOQTRVAQPAFQVALLHRL
LTDGVHITPHYVAGHSGEITAAHLAGILITDATTLLTORATLMTMPBGTWTLHT
TPHSHHTHHTHLLTAHENDLALAIANTPTSLVISGTPHTVQHIITLCOQQGIKTKLPTNHA
FHSHTHNTNLNQLHQTALTYHPHTPLTITANTPPDQLTTPHYTQOAKRTVDTYAT
TQTLHQHGVTYYYIELGPDNTLTLTHLNPPTTTLTTPHHHPQTHLLT
TTPHHHPQTHHTHNTNLNQLHQTALTYHPHTPLTITANTPPDQLTTPHYTQOAKRTVDTYAT
TQTLHQHGVTYYYIELGPDNTLTLTHLNPPTTTLTTPHHHPQTHLLT

[illegible]

and polyketide synthase gene from *Streptomyces avermitilis*

Unpublished
2 (bases 1 to 11096)
Hong, Y.-S. and Lee, J. J.
Direct Submission
Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
Research Institute of Bioscience and Biotechnology, P.O. Box 116,
Yusong-Gu, Taejeon 305-600, South Korea

Location/Qualifiers
1. 11096
/organism="Streptomyces avermitilis"
/strain="ATCC31271"
/db_xref="ATCC:31271"
/db_xref="taxon:33903"
220..>11096
/note="Avr1"
/codon_start=1
/transl_table=1
/product="avermectin polyketide synthase"
/protein_id="AAG09812.1"
/db_xref="GI:9964076"

CDs

FEATURES
source

REFERENCE
AUTHORS
TITLE
JOURNAL

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

BASE COUNT 1756 a 4059 c 3528 g 1753 t
ORIGIN

Query Match 100.0%; Score 246; DB 1; Length 11096;
Best Local Similarity 100.0%; Pred. No. 5.8e-42;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGCATTCGTCGAGCTCGTACGAGCCAT 60
DB 6055 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGCATTCGTCGAGCTCGTACGAGCCAT 6114

QY 61 GTGGCGGAGTCGTCGGCGACCGCGACCCGAAAGCCATCGCGCCGACGAGTCTGTCGCT 120
DB 6115 GTGGCGGAGTCGTCGGCGACCGCGACCCGAAAGCCATCGCGCCGACGAGTCTGTCGCT 6174

QY 121 GCACCTGGCTTCGATTCACCTACGCGCGCTCGAGTTCGAAACCTGCTGATCAAGGCAACA 180
DB 6175 GCACCTGGCTTCGATTCACCTACGCGCGCTCGAGTTCGAAACCTGCTGATCAAGGCAACA 6234

QY 181 GCACCTGGCTTCGATTCACCTACGCGCGCTCGAGTTCGAAACCTGCTGATCAAGGCAACA 240
DB 6235 GCACCTGGCTTCGATTCACCTACGCGCGCTCGAGTTCGAAACCTGCTGATCAAGGCAACA 6294

QY 241 CACCTG 246
DB 6295 CACCTG 6300

RESULT 2
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS Streptomyces avermitilis.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger, J., Leadlay, P.F., Staunton, J., McArthur, H.A. and
Stutzman-Engwall, K.J.
TITLE Polyketides, their preparation, and materials for use therein
JOURNAL Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PEIZER (US); MCARTHUR
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES
source 1. 12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

BASE COUNT 1884 a 4561 c 4005 g 1931 t
ORIGIN

Query Match 100.0%; Score 246; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGCATTCGTCGAGCTCGTACGAGCCAT 60
DB 6752 CTGGCCCATCAGACGTCGGCCGACCAACGGCCGCGCATTCGTCGAGCTCGTACGAGCCAT 6811

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 477.19 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180

Perfect score: 246

Sequence: 1 ctggcccatcagactcggc.....ccaaactcgctacacctg 246

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_on:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pt:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_on:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pi:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	246	100.0	11096	1	AF275943	AF275943 Streptomy
2	246	100.0	12381	6	AX006889	AX006889 Sequence
3	246	100.0	30690	6	E38020	E38020 Avermectin
4	246	100.0	64957	1	AB032367	AB032367 Streptomy
C 5	123.4	50.2	24225	1	SC2C4	AL512302 Streptomy
C 6	113.4	46.1	104326	1	AB070940	AB070940 Streptomy
7	112.2	45.6	20394	1	SN132222	AJ132222 Streptomy
8	112.2	45.6	20394	6	AX067996	AX067996 Sequence
C 9	112.2	45.6	84985	1	SN1278573	AJ278573 Streptomy
10	111.6	45.4	104326	1	AB070940	AB070940 Streptomy
11	110.8	45.0	78210	1	AB070949	AB070949 Streptomy
12	110.2	44.8	11480	1	STMPKS3ORF	L09654 Streptomyce
13	110.2	44.8	24568	1	AF220951	AF220951 Streptomy
14	110.2	44.8	50937	6	AR159871	AR159871 Sequence
15	110	44.7	123580	1	AF263912	AF263912 Streptomy
16	110	44.7	125401	6	AX211739	AX211739 Sequence
17	107.4	43.7	31422	6	E38021	E38021 Avermectin
C 18	107.4	43.7	64957	1	AB032367	AB032367 Streptomy
19	106.2	43.2	113193	1	AF357202	AF357202 Streptomy
20	105.2	42.8	65140	6	AX211705	AX211705 Sequence
21	103.6	42.1	1681	6	AR173228	AR173228 Sequence
22	103.6	42.1	37948	1	AF079138	AF079138 Streptomy
23	102	41.5	39314	1	SGR300302	AJ300302 Streptomy
24	102	41.5	43280	1	SFU78289	U78289 Streptomyce
25	102	41.5	53784	1	AMW23012	AJ23012 Amycolato
26	102	41.5	53789	6	A69720	A69720 Sequence 3
27	102	41.5	90445	1	AF040570	AF040570 Amycolato
28	101	41.1	30000	6	AX250263	AX250263 Sequence
29	101	41.1	32870	1	AF007101	AF007101 Streptomy
30	101	41.1	77534	1	AF235504	AF235504 Streptomy
31	100.4	40.8	2700	6	AX211733	AX211733 Sequence
32	100.4	40.8	2706	1	AF420594	AF420594 Streptomy
33	100.4	40.8	9513	6	AX089460	AX089460 Sequence
34	100.4	40.8	41782	1	AY118081	AY118081 Streptomy
35	100.4	40.8	50000	6	AX089417	AX089417 Sequence
36	100.4	40.8	50000	6	AX089420	AX089420 Sequence
37	100.4	40.8	80161	1	AY007564	AY007564 Saccharop
38	100.4	40.8	80161	6	AR165018	AR165018 Sequence
39	99.4	40.4	30000	6	AX250262	AX250262 Sequence
40	98.8	40.2	5676	6	A69718	A69718 Sequence 1
41	98.8	40.2	41097	1	AF016585	AF016585 Streptomy
C 42	97.8	39.8	1434	6	AR173224	AR173224 Sequence
C 43	97.6	39.7	77534	1	AF235504	AF235504 Streptomy
C 44	97.2	39.5	30000	6	AX250263	AX250263 Sequence
45	96.8	39.3	6459	6	AX089458	AX089458 Sequence

ALIGNMENTS

RESULT 1
AF275943
LOCUS AF275943 11096 bp DNA linear BCT 02-SEP-2000
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene,
partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene


```

RESULT 10
BG786340/c
LOCUS
DEFINITION
SEAUMC006297 Sea urchin primary mesenchyme cell cDNA library,
Strongylocentrotus purpuratus cDNA clone PC_0028_A2_H09_MR 5', mRNA
sequence.
ACCESSION
BG786340
VERSION
BG786340.1 GI:14157353
KEYWORDS
EST.
SOURCE
Strongylocentrotus purpuratus.
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 698)
AUTHORS
Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
Ettensohn,C.A.
TITLE
A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
JOURNAL
Development 128 (13), 2615-2627 (2001)
MEDLINE
21384984
COMMENT
Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
FEATURES
source
Location/Qualifiers
1..698
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC 0028 A2 H09 MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; oligo
dt priming from poly A+ RNA, directionally cloned"
BASE COUNT
9 a 0 c 596 g 53 t 30 others
ORIGIN
Query Match 29.2%; Score 155.8; DB 12; Length 688;
Best Local Similarity 55.7%; Pred. No. 8.6e-17;
Matches 291; Conservative 0; Mismatches 227; Indels 4; Gaps 1;

Qy 5 CCACCTCTATCAGCGGGGAACCGGGCGCTCTGCGACCCCTCACCCTCACCACCTCACC 64
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 CCCCCCCCCCACCACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 507

Qy 65 CCCACCAACCCACCCACACCTCTCTCTCACCAGCGAGCGGCCGCCACACCGCCCAAG 124
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
506 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 447

Qy 125 CACAACACCTCACCACCCCACTCAACAAAGGAGGATCCACTCACCATCACCACCTG-- 182
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
446 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 387

Qy 183 --CGACACAGCAACCCAGACCAACTCCAACTCTCTAACACATCTCCCCCAACAAC 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 CCCCCCCCCCACCACCCCAACACACANCCCCCCCCCCCCCCCCCCCCCCCCCCCC 327

Qy 241 CCCCTCACCACGGTCATCCACCGCAGGATCTCTCGACGAGCGACCTCACCACCTC 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
326 CCCACCAAAACACACCCCCCCCCAGNCCACCCCAACCCCAACNCCNCCNCCNCCC 267

Qy 301 ACCCCCACCCCACTCAACAGTCTCTCGGGCCAAAGCCACAGCGCCCACTCTCTCAC 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 CCCCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 207

Qy 361 CAATCTACCCNACACACCCCTCTCAGCGCTTGTCTCTCTCTCTCTCTCTCTCTCT 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 206 CACCCACCCACACANCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 147
Qy 421 TTTCGGCGCAGCCGCGCAAGCAACTACGCGCAGCAAGCCTTACCTCGAGCCCTCGCC 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
146 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 87
Qy 481 CACCCACCGCCACACACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 522
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 CNAACACNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 45

RESULT 11
CNS046XY
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
086D07 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL277279.1 GI:8011482
VERSION
GSS; genome survey sequence.
KEYWORDS
Tetraodon nigroviridis.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 638)
AUTHORS
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 638)
AUTHORS
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
Unpublished
REFERENCE
3 (bases 1 to 638)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (12-APR-2000)
COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
Location/Qualifiers
1..638
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="086D07"
/clone_lib="G"
/notes="Genoscope sequence ID : COBG086CB04SP1-end ;
PUC-Ori"
BASE COUNT
59 a 513 c 16 g 21 t 29 others
ORIGIN
Query Match 29.1%; Score 155.6; DB 17; Length 638;
Best Local Similarity 52.9%; Pred. No. 9.3e-17;
Matches 275; Conservative 23; Mismatches 222; Indels 0; Gaps 0;

Qy 3 AACCCCTCTATCACCAGCGGGAACCGGGCGCTCTGCGACCCCTCACCACCTCACC 62
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 ATCAGACCCACACACCCCAACCCCGCTCTCTCACCACCCCAACCTGSMCMCAACMCCC 80

Qy 63 CACCCACCAACCCACCCACACCTCTCTCTCAGCGGAGCGGCGCCCAACCCCCCA 122
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
81 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 140

Qy 123 CGCACAAACCTCACCACCCCACTCCAAACAAAAGGAGGATCCACTCACCATCACC 182
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1986 row: k column: 12
High quality sequence stop: 218.

FEATURES

Location/Qualifiers
1. 1161
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5474051"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 53 a 84 c 107 g 88 t 69 others
ORIGIN

Query Match 29.3%; Score 156.6; DB 14; Length 1161;
Best Local Similarity 54.4%; Pred. No. 6.2e-17;
Matches 282; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
Qy 5 CCACCTCTATCAGCGGGAACCGGCGCTCTGCGACCCACTCACCACCACTCACCACCA 64
Db 312 CCCCCCCCCCCCCNNNNCC 371
Qy 65 CCCACCAACCCAGCAACCTCTCTCACCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124
Db 372 CC 431
Qy 125 CACCAACCTCACCACCACTCTCAACAAAGGATCCACTCACCATCACCACCTGCG 184
Db 432 CCCCNGCC 491
Qy 185 ACACGAGCAACCCAGCAACCTCTCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 244
Db 492 CC 551
Qy 245 TCACCAAGCTATCAGCAACGATCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 304
Db 552 CC 611
Qy 305 CCACCAACTCAACAACTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 364
Db 612 CC 671
Qy 365 TCACCAACACACCCCTCAGCGCTTCTGCTCTACTCTCTGCGGCGGCGGCGGCGGCGG 424
Db 672 NNCCT 731
Qy 425 GGCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 484
Db 732 NCCC 791
Qy 485 ACCGCGCACACCACT 522
Db 792 CC 829

RESULT 9

BI416700/c 842 bp mRNA linear EST 15-AUG-2001
LOCUS
DEFINITION
hasp001xp03f Heterobasidion annosum - Scots pine infection stage
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp001xp03f, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
BI416700
BI416700.1 GI:15187723
EST.

SOURCE

ORGANISM
Pinus sylvestris/Heterobasidion annosum.

REFERENCE

1 (bases 1 to 842)
Asiegbu.F.O., Nahalkova.J., Choi.W., Stenlid.J. and Dean.R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)

JOURNAL

Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish
University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden

AUTHORS

Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se

Seq primer:

T7 primer.
Location/Qualifiers

FEATURES

1. 842
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xp03f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."
BASE COUNT 17 a 14 c 654 g 50 t 107 others
ORIGIN

Query Match 29.3%; Score 156.4; DB 13; Length 842;
Best Local Similarity 54.1%; Pred. No. 6.8e-17;
Matches 280; Conservative 0; Mismatches 238; Indels 0; Gaps 0;
Qy 5 CCACCTCTATCAGCGGGAACCGGCGCTCTGCGACCCACTCACCACCACTCACCACCA 64
Db 730 CCCCCCCCCACACC 671
Qy 65 CCCACCAACCCAGCAACCTCTCTCACCAGCGGAGCGGCGGCGGCGGCGGCGGCGG 124
Db 670 ACCCAACCCCGCC 611
Qy 125 CACCAACCTCACCACCACTCTCAACAAAGGATCCACTCACCATCACCACCTGCG 184
Db 610 CC 551
Qy 185 ACACGAGCAACCCAGCAACCTCTCAACAACTCTCTCAACAGCATCCCCCGGCGGCGG 244
Db 550 CACCCACCCCGACCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 491
Qy 245 TCACCAAGCTATCAGCAACGATCTCTGAGAGGAGGCGGCGGCGGCGGCGGCGGCGG 304
Db 490 CCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 431
Qy 305 CCACCAACTCAACAACTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 364
Db 430 CC 371
Qy 365 TCACCAACACACCCCTCAGCGCTTCTGCTCTACTCTCTCTCTCTCTCTCTCTCTCT 424
Db 370 CC 311
Qy 425 GCGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 484
Db 310 CCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 251
Qy 485 ACCGCGCACACCACT 522
Db 250 CCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 213


```

Qy 305 CCACCAACTCAAGAGTCTCGCGCCGAAGCCACAGCGGCACCTCTCCACCAAC 364
Db 467 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 408
Qy 365 TCACCCCAACACACACCCCCCTACCGCCCTTGGTCTCTACTCTCCGCGCGCGCCACCTTCG 424
Db 407 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 348
Qy 425 GCGCACCGGCGCAAGCCAACTACGCGGAGGCAAGCGCTACTCGAGCGCCCTCGCCCAAC 484
Db 347 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 288
Qy 485 ACCGCGACACCCACCACTCTCCCGCGCCACGAGCATCGCC 522
Db 287 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 250

RESULT 5
BI416659/c
LOCUS BI416659 806 bp mRNA linear EST 15-AUG-2001
DEFINITION haep001xm18f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep) Pinus sylvestris/Heterobasidion annosum cDNA clone haep001xm18f, mRNA sequence.
ACCESSION BI416659
VERSION BI416659.1 GI:15187682
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 806)
AUTHORS Asiegbu, P.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the intersection of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
source
1..806
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="haep001xm18f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep)"
/dev_stages="Seedling roots of scots pine were infected for 6 days with H. annosum"
/notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."
BASE COUNT 13 a 15 c 676 g 35 t 67 others
ORIGIN
Query Match 29.6%; Score 158; DB 13; Length 806;
Best Local Similarity 56.0%; Pred. No. 3.7e-17;
Matches 290; Conservative 0; Mismatches 228; Indels 0; Gaps 0;
Qy 5 CCACCTCATCAGCGGCGAAGCCGCGCTCTCTCCTCACCAGCGGACCGGCCCCCACCACCTCACC 64
Db 707 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 648
Qy 65 CCCACCAACCCACCAACCACTCTCTCTCCTCACCAGCGGACCGGCCCCCACCACCTCACC 124
Db 647 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 588
Qy 125 CACAACACCTCACCACCCAACTCCAAACAAAAGGATCCACCTCACCATCACCACCTCGC 184

```

```

Db 587 CCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 528
Qy 185 ACACGAGCAACACAGACCAACTCCAAACAATCTCTCAAGACCATCTCCGCGCCACACACCCC 244
Db 527 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 468
Qy 245 TCACGACCGGTATCCACACGCGAGGATCTCTGACGAGCGCACCTTCACCAACTCACC 304
Db 467 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 408
Qy 305 CCACCCAACTCAAAACAGTCTCTCCGCGCCAAAGCCACAGCGCCACCTCTCCACCAAC 364
Db 407 CCCCCCACCGCCGCCNCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 348
Qy 365 TCACCCCAACACACCGCCCTCAGCGCTTGGTCTCTACTCTCTCGGCGCGGCGACCTTCG 424
Db 347 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 288
Qy 425 GCGCACCGGCGCAAGCCAACTACGCGCGAGCAAGCGCTACTCTCGAGCGCCCTCGCCCAAC 484
Db 287 CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 228
Qy 485 ACCGCGCACACCCACCACTCTCCCGCGCCACGAGCATCGCC 522
Db 227 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 190

RESULT 6
AL570062/c
LOCUS AL570062 715 bp mRNA linear EST 16-FEB-2001
DEFINITION AL570062 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1006YC11 3 prime, mRNA sequence.
ACCESSION AL570062
VERSION AL570062.1 GI:12926016
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 715)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1006YC11"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 35 a 31 c 574 g 37 t 38 others
ORIGIN
Query Match 29.4%; Score 157; DB 9; Length 715;
Best Local Similarity 52.7%; Pred. No. 5.4e-17;
Matches 280; Conservative 23; Mismatches 228; Indels 0; Gaps 0;
Qy 4 ACCACCTCATCAGCGGCGAAGCCGCGGCGCTCGCCACCCACCTTCACCCACCACTCACC 63

```

[illegible]

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5480305"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 73 a 824 c 62 g 234 t 43 others
ORIGIN

Query Match 30.0%; Score 160; DB 14; Length 1236;
Best Local Similarity 54.8%; Pred. No. 1.7e-17;
Matches 289; Conservative 0; Mismatches 238; Indels 0; Gaps 0;
Qy 5 CCACCTCATCACCGGGAACCGGCGCTCTCTCAGCAGCGGCGCCACCACTCACCACTCACCA 64
Db 184 CCCCCCTCCCCCNC 243
Qy 65 CCCACCAACCCACCAACCACTCTCTCTCAGCAGCGGCGCCACCACTCACCACTCACCA 124
Db 244 CCCCCCNC 303
Qy 125 CACAACACTTACCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 184
Db 304 CCCCCCNC 363
Qy 185 ACACGAGAACCCAGCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 244
Db 364 CCCCCCNC 423
Qy 245 TCACCACTGTCATCACACCGAGGATCTCTCAGCAGCGGCGCCACCACTCACCACTCAC 304
Db 424 CCCCCCNC 483
Qy 305 CCACCCAACTCAACAGCTCTCTCGGCGCAAGGCGCACAGGCGGCGGCGGCGGCGGCG 364
Db 484 CCCCCCNC 543
Qy 365 TCACCAACACACCCCTTCT 424
Db 544 CCCCCCNC 603
Qy 425 GCGCACCGGCGAAGCACTACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 484
Db 604 CCCCCCNC 663
Qy 485 ACCGCGACACCACT 531
Db 664 CCCCCCNC 710

RESULT 2
BM913931/c 1243 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT 6608063 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478793
5', mRNA sequence.
ACCESSION BM913931
VERSION BM913931.1 GI:19364310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1243)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LCM1999 row: a column: 02
High quality sequence stop: 369.

FEATURES

source

1. 1243

Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5478793"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 47 a 193 c 776 g 116 t 111 others
ORIGIN

Query Match 30.0%; Score 160; DB 14; Length 1243;

Best Local Similarity 53.5%; Pred. No. 1.7e-17;

Matches 277; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

Qy 5 CCACCTCATCACCGGGAACCGGCGCTCTCTCAGCAGCGGCGCCACCACTCACCACTCACCA 64
Db 954 CCCCCCNC 895
Qy 65 CCCACCAACCCACCAACCACTCTCTCTCAGCAGCGGCGCCACCACTCACCACTCACCA 124
Db 894 CCCCCCNC 835
Qy 125 CACAACACTTACCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 184
Db 834 CCCCCCNC 775
Qy 185 ACACGAGAACCCAGCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 244
Db 774 CCCCCCNC 715
Qy 245 TCACCACTGTCATCACACCGAGGATCTCTCAGCAGCGGCGCCACCACTCACCACTCAC 304
Db 714 CCCCCCNC 655
Qy 305 CCACCCAACTCAACAGCTCTCTCGGCGCAAGGCGCACAGGCGGCGGCGGCGGCGGCG 364
Db 654 CCCCCCNC 595
Qy 365 TCACCCAACTCAACAGCT 424
Db 594 CCCCCCNC 535
Qy 425 GCGCACCGGCGAAGCACTACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 484
Db 534 CCCCCCNC 475
Qy 485 ACCGCGACACCACT 522
Db 474 CCCCCCNC 437

RESULT 3

B1416575/c

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 684.535 Seconds
(without alignments)

12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676

Perfect score: 534

Sequence: 1 ggaaccacctcatcaccgg.....gcatacgtggggcactcg 534

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160	30.0	1236	14	BM914619
C 2	160	30.0	1243	14	BM913931
C 3	159.6	29.9	931	13	BI416575
C 4	159.2	29.8	1015	13	BI416650
C 5	158	29.6	806	13	BI416659
C 6	157	29.4	715	9	AL570062
					AL570062 AL570062

7	156.6	29.3	897	17	AZ186654	AZ186654 SP_1007_B
8	156.6	29.3	1161	14	BM912505	BM912505 AGENCOURT
C 9	156.4	29.3	842	13	BI416700	BI416700 hasp001xp
C 10	155.8	29.2	688	12	BG786340	BG786340 SEAUMC006
C 11	155.6	29.1	638	17	CNS046XY	AL277279 Tetraodon
C 12	155.6	29.1	1089	14	BQ939440	BQ939440 AGENCOURT
C 13	155.4	29.1	1228	12	BG067997	BG067997 H3060F02-
C 14	155.4	29.1	1303	14	BM908878	BM908878 AGENCOURT
C 15	155	29.0	1197	13	BI416470	BI416470 hasp001xa
C 16	154.8	29.0	660	17	AG089410	AG089410 Pan trogl
C 17	154.8	29.0	1005	17	BI418610	BI418610 ENTP267TR
C 18	154.6	29.0	703	17	AZ185816	AZ185816 SP_1006_A
C 19	154.6	29.0	1011	17	AG062236	AG062236 Pan trogl
C 20	154.6	29.0	1089	14	BQ930965	BQ930965 AGENCOURT
C 21	154.6	29.0	1138	17	AG152606	AG152606 Pan trogl
C 22	154.4	28.9	703	12	BG786294	BG786294 SEAUMC006
C 23	154.4	28.9	863	13	BI416699	BI416699 hasp001xp
C 24	154	28.8	748	17	AZ187748	AZ187748 SP_1009_B
C 25	154	28.8	765	17	AG043737	AG043737 Pan trogl
C 26	154	28.8	843	17	AZ200896	AZ200896 SP_1011_A
C 27	153.8	28.8	1047	14	BQ894173	BQ894173 AGENCOURT
C 28	153.6	28.8	1088	12	BF256580	BF256580 HVSMEF001
C 29	153.2	28.7	877	17	AZ185000	AZ185000 SP_1004_A
C 30	153.2	28.7	993	17	AG081720	AG081720 Pan trogl
C 31	153	28.7	1184	14	BM911719	BM911719 AGENCOURT
C 32	152.8	28.6	903	17	BI436778	BI436778 ENTPA49TF
C 33	152.8	28.6	1089	13	BM468944	BM468944 AGENCOURT
C 34	152.6	28.6	817	17	AZ196702	AZ196702 SP_1032_B
C 35	152.6	28.6	897	13	BI416517	BI416517 hasp001xd
C 36	152.4	28.5	707	17	AZ193616	AZ193616 SP_1023_B
C 37	152.4	28.5	780	10	BE641449	BE641449 Cr12_3_E2
C 38	152.4	28.5	1222	17	BI446899	BI446899 ENTQ47TF
C 39	152.2	28.5	803	13	BI416474	BI416474 hasp001xa
C 40	152.2	28.5	993	17	AG033032	AG033032 Pan trogl
C 41	152	28.5	672	17	AG133158	AG133158 Pan trogl
C 42	152	28.5	888	17	AZ189327	AZ189327 SP_1014_A
C 43	152	28.5	1052	17	AQ740804	AQ740804 HS_5508_A
C 44	152	28.5	1478	14	BM911600	BM911600 AGENCOURT
C 45	151.8	28.4	540	17	CNS0103K	AL098474 Drosophila

ALIGNMENTS

RESULT 1
LOCUS BM914619
DEFINITION AGENCOURT_6615286 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:5480305
5', mRNA sequence.
ACCESSION BM914619
VERSION BM914619.1
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1236)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Mark Watson

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLCM2002 row: p column: 02

High quality sequence stop: 147.

Location/Qualifiers

1. 1236

source


```
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Query Match      36.8%; Score 196.4; DB 9; Length 15872;
Best Local Similarity 60.5%; Pred. No. 6.6e-35;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY      1  GGAACCACTCTATACCGGCGGAACCGCGCCCTCGCCACCCACCTCACCCACACACCTC 60
Db      7337  GGCACCGTCTCTGTACCGGGGACCGGAGGCTCGGAGCGTCTCGCCGCCACCTC 7396

QY      61  ACCACCCACCAACCCACCAACCTCTCTCTACAGCGGAAACCGGCCCCCACCACCCCC 120
Db      7397  GTGAGCGTCAACGGGTGCGCGGCTGTCTGTGTGAGCGCGCGGGGCGGAGGCCCC 7456

QY      121  CAGGACACACTCTACACCCCACTCAACAAAGGATCCACCTACCATCACCAACC 180
Db      7457  GCGCGCGCGGACCTGGGCGAGGACCTCGCGGGCTCGGGCGGAGGTGGCGTTTCGCGCC 7516

QY      181  TCGGACACGAGCAACCCAGACCACTCAACAACTCTCAACACCATCCCCCACAACAC 240
Db      7517  GCGAGCGCGGACCGCGGAGCTTGGCGGGGCGATCGCCACCGTCCCGCGGAGCAT 7576

QY      241  CCCCTCACCACTCATTCACACCGGAGGATCTCTGAGAGCGCCACCTCACCAACCTC 300
Db      7577  CCGCTGACGGCGTCTGTGACACGCGGGAGTGTGACGACGCGAGCGGTGGAGGGGCTC 7636

QY      301  ACCCCACCCCACTCAACAGTCTCTCGGGCCAAAGCCACAGCGGCCACCTCTCTCAC 360
Db      7637  ACACCGGAACGGCTGAGCGGTACTCGGCCCGAAGTCTGAGCGCGCGTGGAACTGAC 7696

QY      361  CAACTCACCAACACACCCCTCTACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db      7697  GAGCTCACCAAGGACTGTGGGCTGAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 7756

QY      421  TTGGGCGCACCGGCGCAAGCCAACTAGCGCGCAGCCACGCTTACTCGAGCCCTTCGCC 480
Db      7757  GTCGGCACCGCGGCGCAGGCCAACTAGCGGGGCGCAACAGCGGGCTCGAGCGCCCTCG 7816

QY      481  CACCACCGGCACACCCACCACTCTCCCGGCCACAGCATGCTTGGGGGACCTGG 534
Db      7817  GCCACCGCGCGCCACCGGCGCTTGGCGGCCACGCTGCTGCGCTGGGGGCTCTGG 7870
```

Search completed: June 18, 2003, 05:08:24
Job time : 66.1981 secs

Qy	61	ACGACCCACAA	CCGACCCCAAC	CACTCTCTCTCA	CAGCGGAA	CCGGCCCCCA	CCGCC	120
Db	9593	GCCACCA	CCGAGCGG	---AACA	CTCTCTCTCGT	CAGCGGAGG	CGGAA	9649
Qy	121	CAGCGA	CAACACT	CACCA	CCCAACT	CCAA	AAAAAGG	180
Db	9650	GGAGCC	ACCAC	CTAC	CGCGGAA	CTAC	CGCAT	9709
Qy	181	TGCGAC	ACCAAG	CCAGAC	CAACT	CCAA	CAACT	240
Db	9710	TGCGA	CGTCCG	ACCCCA	CGCAT	CGCAC	CTCTCGA	9769
Qy	241	CCCTT	CACCA	CGTCA	TCAC	ACCG	CAGGCA	300
Db	9770	CCCTT	CACCG	CGTCTG	TCAC	ACCG	CGCGCG	9829
Qy	301	ACGCCC	ACCA	CTCA	CAAC	GTCT	CCGCGC	360
Db	9830	ACCGCG	AGCAG	GTCCG	GCGGG	CCAC	CGGTG	9889
Qy	361	CAACT	CACCA	CA	ACCCCC	CTCA	CGCGT	420
Db	9890	GAGT	CACCG	GACCT	CGAC	CTCG	ACCTCG	9949
Qy	421	TTGCGG	CGCAC	CCCGG	CAAG	CAACT	ACGCG	480
Db	9950	CTGGC	ATCCCC	CGT	CAGG	CAACT	AGCCCC	10009
Qy	481	CACCA	CGGC	ACA	CCCA	CACT	CCCG	534
Db	10010	GCT	CGCG	CGCG	GCAC	ACCG	CGCG	10063

RESULT 9

```

US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

```

Query Match 41.9%; Score 223.6; DB 9; Length 38506;
Best Local Similarity 65.0%; Pred. No. 6.6e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
QY 1 GGAACCACTCTATACCGGGGAACCGGGGCCCTCGCCACCACTCACCCACACGCTC 60

Db	6505	GGCACCGTCTCTATACCCGCGCGCATCGGAGCCCTCGGAGCCACCGCGACGCTGGATG	6564
Qy	61	ACCACCCACCAACCCCAACACCTCTCTCTCACAGCGCAACCGGCGCCCCACACCCCC	120
Db	6565	GCCACACGGAGCG--AACACCTCTCTCTGTCAGCGCGAGCGGCAACGCCCCC	6621
Qy	121	CAGGCACACACTGACACCCAACTCCAAACAAAGGAGTCACTTCAACCATCACCAC	180
Db	6622	GGAGCCACCCAACTCAACCGCGAACTCAGCGCATCGGCGCGCGCTCAACATGCGCGCC	6681
Qy	181	TGGCACACAGCAACCCAGACCAACTCCAACAACCTCTCAACACCATCCCCCACAACAC	240
Db	6682	TGGCAGTGCAGACCCCAACGGCATTGGCACCTCTCGACGCCATCCCGCGGAGAG	6741
Qy	241	CCCTTCACACCTGTATACACACCGGAGGATCTTCGACGACGCCACCTTCAACCACTC	300
Db	6742	CCCTCACCCCGCTGCTCCACACCGCGCGCGCTCGAGACGGCATCTGTGGACACGCTG	6801
Qy	301	ACCCCAACCCAACTCAACAACGTCCTCCGCGCCAAAGCCACAGCGGCCCACTTCCTCCAC	360
Db	6802	ACCGCGAGGAGTCCGGGGGCCACAGTGCAGAGGCCGTGGCGCCTTCGTGCTCGAC	6861
Qy	361	CAACTCACCAACACACCCCTTCAACCGCTTGTCTCTACTCTTCGCGCGCGCCACC	420
Db	6862	GAGCTGACCGGACCTCGACCTCGAGCGGTTGTTGTTCTTCTCGTCCGTTCGAGCACT	6921
Qy	421	TTTCGGCGCACCGGGCCAACTACGCGGAGCCAAAGCTTACTTCAAGCGCCCTCGCC	480
Db	6922	CTGGGCATCCCGGTGAGGGCAACTAGCCCCCGCAACAGCCTTACTCGAGCGCCTCGCG	6981
Qy	481	CACACCGCCACACCCACACCTTCCCGGCCACAGCATCGCGTGGGACCTGG	534
Db	6982	GCTTCGCGCGCGGCACCGCGCGGTTCGCGCTCTCGTTGGCTTGGGACCGTGG	7035

RESULT, T 10

```

US-09-860-846-34
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-34

```

	Query Match	40.1%;	Score 214;	DB 9;	Length 4689;
	Best Local Similarity	63.9%;	Pred. No. 9.1e-39;		
	Matches 341; Conservative	0;	Mismatches 190;	Indels 3;	Gaps 1;
Qy	1	GGAAACCACCTCATCACCGGCGGAACGGCGCCCTTGCACCACTATCCCACACACTTC	60		
Dd	3346	GGCACCCTCTCATCACCGGCGGACCGAGCCCTCGGAGCCACGCCGACGCTGGATG	3405		
Qy	61	ACCACCCACAACCCACCAACTCTCTCTTCAACAGCGCAACGGGCCCCACACCCCC	120		
Dd	3406	GCCCCACCCGGAGCCG---AACACCTCTCTCTCTGTGAGCGGAGGGCGAACAAGGCCCC	3462		
Qy	121	CAGGCAACAACACTCATACCAACCAACTCCAACAAAAAGGCATCCACCTCAACCATCAACCAC	180		

;; PRIOR APPLICATION NUMBER: 09/105,537
;; PRIOR FILING DATE: 1998-06-26
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5
;; LENGTH: 36778
;; TYPE: DNA
;; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 41.9%; Score 223.6; DB 9; Length 36778;
Best Local Similarity 65.0%; Pred. No. 6.6e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
1 GGAACCAACCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 60
8363 GGCACCGTCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 8422
61 ACCACCAACCAACCAACCAACCAACCTCTCTTCCAGCGGGAACCGCGCCCAACCCACCC 120
8423 GCCACCAACCGGAGCGG---AACAACCTCTCTCTCTCGTCAGCGCGGAGGCGGAACAAGCC 8479
121 CAGGCAACACCTTACCAACCAACCTCAACCAACCAACCAACCAACCAACCAACCAAC 180
8480 GGAGCAACCAACCTTACCAACCAACCTCAACCAACCAACCAACCAACCAACCAACCA 8539
181 TCGGACCAACCAACCAACCAACCAACCTCAACCAACCAACCTCAACCAACCAACCAAC 240
8540 TGGAGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 8599
241 CCCCTCACCACCGTCTATCAGCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTC 300
8600 CCCCTCACCACCGTCTATCAGCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTC 8659
301 ACCCCACCAACCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCC 360
8660 ACCCGGAGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAGG 8719
361 CAACTCACCACCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCC 420
8720 GAGCTGACCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAG 480
421 TTGCGGCAACCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCT 534
8780 CTGGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAGGACCAACCT 8839
481 CACCACCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCT 534
8840 GCTCGCGCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCT 8893

RESULT 7
US-09-861-289-5
;; Sequence 5, Application US/09861289
;; Patent No. US20020110897A1
;; GENERAL INFORMATION:
;; APPLICANT: Sherman, D.H.
;; APPLICANT: Liu, H.
;; APPLICANT: Xue, Y.
;; APPLICANT: Zhao, L.
;; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;; FILE REFERENCE: 600.438US1
;; CURRENT APPLICATION NUMBER: US/09/861,289
;; PRIOR FILING DATE: 2001-05-18
;; PRIOR APPLICATION NUMBER: 09/105,537
;; PRIOR FILING DATE: 1998-06-26
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5
;; LENGTH: 36778
;; TYPE: DNA
;; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 41.9%; Score 223.6; DB 10; Length 36778;
Best Local Similarity 65.0%; Pred. No. 6.6e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
1 GGAACCAACCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 60
8363 GGCACCGTCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 8422
61 ACCACCAACCAACCAACCAACCAACCTCTCTTCCAGCGGGAACCGCGCCCAACCCACCC 120
8423 GCCACCAACCGGAGCGG---AACAACCTCTCTCTCTCGTCAGCGCGGAGGCGGAACAAGCC 8479
121 CAGGCAACACCTTACCAACCAACCTCAACCAACCAACCAACCAACCAACCAACCAAC 180
8480 GGAGCAACCAACCTTACCAACCAACCTCAACCAACCAACCAACCAACCAACCAACCA 8539
181 TCGGACCAACCAACCAACCAACCAACCTCAACCAACCAACCTCAACCAACCAACCAAC 240
8540 TGGAGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 8599
241 CCCCTCACCACCGTCTATCAGCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTC 300
8600 CCCCTCACCACCGTCTATCAGCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTC 8659
301 ACCCCACCAACCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCC 360
8660 ACCCGGAGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAGG 8719
361 CAACTCACCACCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCC 420
8720 GAGCTGACCGGAGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAG 480
421 TTGCGGCAACCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCT 534
8780 CTGGGATCTCTCGAGGAGGACCAACCTCAACCAACCTCTCTCGAGGAGGACCAACCT 8839
481 CACCACCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCTCT 534
8840 GCTCGCGCGGCAACCAACCAACCTCTCTCGAGGAGGACCAACCTCAACCAACCTCT 8893

RESULT 8
US-09-988-384B-5
;; Sequence 5, Application US/09988384B
;; Publication No. US20030073824A1
;; GENERAL INFORMATION:
;; APPLICANT: Sherman, D.H.
;; APPLICANT: Liu, H.
;; APPLICANT: Xue, Y.
;; APPLICANT: Zhao, L.
;; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;; FILE REFERENCE: 600.536US1
;; CURRENT APPLICATION NUMBER: US/09/988,384B
;; PRIOR FILING DATE: 2001-11-19
;; PRIOR APPLICATION NUMBER: PCT/US99/14398
;; PRIOR FILING DATE: 1999-06-25
;; PRIOR APPLICATION NUMBER: US 09/105,537
;; PRIOR FILING DATE: 1998-06-26
;; NUMBER OF SEQ ID NOS: 53
;; SEQ ID NO 5
;; LENGTH: 37948
;; TYPE: DNA
;; ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match 41.9%; Score 223.6; DB 9; Length 37948;
Best Local Similarity 65.0%; Pred. No. 6.6e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;
1 GGAACCAACCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 60
9533 GGCACCGTCTCATACCGCGGGAACCGCGCCCTCGCCACCCACCTCAACCCACCACTC 9592

QY 61 ACCACCCACCAACCCACCAACCTCTCTCACCAGCGGACCGGCCCCACACCCCC 120
Db 13516 GTGGGGAACACAGCGCCAAACCTCTCTCACCCTCAGAGAGGCGCGGTCTCG 13575
QY 121 CACGCAACACCTCTACCAACCACTCTCAACAAAGGATCCACCTCACCATCACCACC 180
Db 13576 GCGCGGAGGCTCTGCAAGCGAGCTCGAAGCGCTGGGGGCTCTCGGTACCCCTCGTGGG 13635
QY 181 TCGGACACCAACCCACCAACCTCTCAACAACTCTCTCAACACCTCTCTCAACAC 240
Db 13636 TCGGACCTGTGCGGACCCACCGCGCTCTCGGACCTCTCTGACAGCATCTCCGAGGGATCAT 13695
QY 241 CCCTCACCACCGTCTATCACAACCGGAGCATCTCTGAGAGCGCCACCTCACCACCTC 300
Db 13696 CCGATCAGCGCGTCTGTCAGCGCGCGGCGCTCGAGAGCGGCGCTCTCGGTAGCATG 13755
QY 301 ACCCCACCACTCAACAAAGTCTCTCGCGGCAAGCCACAGCGGCCACCTCTCTCCAC 360
Db 13756 AGCGCGAGCGATCGCTCGCGCTCTTGTACCCCAAGCTCGATCGATCGCTCTCGGCGTC 13815
QY 361 CAACTCACCACACACCCCTCTCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 13816 GAGCTCAGCGCGTCTGTCAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13875
QY 421 TTGCGGCGCACCCGCGCAAGCAACTACCGCGAGCGCAAGCTCTCTCTCTCTCTCTCTCT 300
Db 13876 CTTGGTGGTCCAGGTCTGATCGAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13935
QY 481 CACACCGCGCACCCACCT 534
Db 13936 CATCACCGCGCGCCCAAGGACT 13989

RESULT 15

US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/729,214
; FILING DATE: TBA
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-729-214-6

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
QY 1 GGAACCAACCTCTATCACCAGCGGAAACCGGCGCTCTGCGACCCACCTCACCACCCACCTC 60
Db 13456 GGGACCGTCTCTATCAGGGAGGACCGGAGCTCTAGGAGTCTTGGTTCGACGCCACCTC 13515
QY 61 ACCACCCACCAACCCACCAACCT 120
Db 13516 GTCGGAAACACAGCGCCAAACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCG 13575
QY 121 CACGCAACACCTCACCACCT 180
Db 13576 GCGCGGAGGCTCTGCGAAGCGAGCTCTGAAAGCGCTCTGCGGCGCTCTCTCTCTCTCTCT 13635
QY 181 TCGGACACCAACCCACCAACCT 240
Db 13636 TCGGACGTGGCGGACCCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13695
QY 241 CCCTCACCACCGTCTATCACAACCGGAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
Db 13696 CCGATCAGCGCGTCTGTCAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13755
QY 301 ACCCCACCACTCAACAAAGTCT 360
Db 13756 AGCGCGAGCGATCGCT 13815
QY 361 CAACTCACCACACACCCCT 420
Db 13816 GAGCTCAGCGCGTCTGTCAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13875
QY 421 TTGCGGCGCACCCGCGCAAGCAACTACCGCGAGCGCAAGCTCTCTCTCTCTCTCTCTCTCTCT 480
Db 13876 CTTGGTGGTCCAGGTCTGATCGAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13935
QY 481 CACACCGCGCACCCACCT 534
Db 13936 CATCACCGCGCGCCCAAGGACT 13989

Search completed: June 18, 2003, 01:06:58
Job time : 27.1323 secs

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ciba-Geigy Corporation
;; STREET: 520 White Plains Road, P.O. Box 2005
;; CITY: Tarrytown
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10591
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/764,233A
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/729,214
;; FILING DATE: 09-OCT-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/258,261
;; FILING DATE: 08-JUN-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241
;; REFERENCE/DOCKET NUMBER: 1506/CIP6
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919) 541-8587
;; TELEFAX: (919) 541-8689
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 28958 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; ORIGINAL SOURCE:
;; ORGANISM: Sorangium cellulosum
;; IMMEDIATE SOURCE:
;; CLONE: p98/1
;; US-08-764-233A-4

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 1 GGACCCACCTCATCACCGCGGACCGGGCGCTCGCCACCGACCTTACCCACCACTTC 60
Db |||||
13456 GGGACCGTCTCATCACGGGAGGACCGGGAGCTCTGCTCGACGCCACCTC 13515

QY 61 ACCACCCACCAACCCACCACTCTCTCCACCGAGCGGCGGCGGCGGCGGCGGCGG 120
Db |||||
13516 GTGCGGAACACAGCGCCAAACCTGCTCTTACCTCGAGGAGGCGGCGGCTCGG 13575

QY 121 CACGCACACCTCATCACCACTCCCAACCAAGGATCCCTCACCATCACCACTC 180
Db |||||
13576 GCGCGGGAGCTCTCGAAGCGAGCTCGAAGCGCTGGGGCGCTCGGTACACCTCTCGG 13635

QY 181 TCGACACAGCAACCCAGACCACTCCAACTCTCTCAACACCATCCCGCCCAACAC 240
Db |||||
13636 TCGGAGTGGCGGACCCACCGCGCCCTCCGAGCCCTCTCGACAGATCCCGAGGATCAT 13695

QY 241 CCCTCACCACTCATCACACCGCGGATCTCGAGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db |||||
13696 CGATCAGGGCGTGTGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13755

QY 301 ACCCCACCACTCAACAACTCTCTCGCGGCAAGCCCAAGCGGCGGCGGCGGCGGCGG 360
Db |||||
13756 AGCGCGGAGCGATCGCTCGGCTCTTGACCCCAAGCTCGATGCGGCTTGATTCAT 13815

QY 361 CAACTACCCCAACACACCCCTCTACCGCTTCTGCTCTTACTCTCTCGCGCGCGGCGG 420
Db |||||
13816 GAGCTCACCCAGGAGCGGTCGCGGCTTCTGCTCTTCTCTCGCGCGGCGGCGGCGG 13875

QY 421 TTCCGCGCACCGCGCAAGCCAACTACGCGGAGCAAGCGCTTACCTCGACGCGCTCGGC 480
Db |||||
13876 CTTGGTGGTCCAGGTCACTGCACTACGCGGCTGCCAATGCTTCTCTCATGGCTCGCA 13935

QY 481 CACCAACCGCCACACCCACACCTCTCCCGGCCACAGCATCGCTGGGGCACCTGG 534
Db |||||
13936 CATCACCGGCGCGCAAGGACTCCGAGCGGCTTCTGCTCGGCGGCTACTGG 13989

RESULT 14
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,335A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-457-335A-6

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 1 GGACCCACCTCATCACCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
Db |||||
13456 GGGACCGTCTCATCACGGGAGGACCGGGAGCTCTGCTCGGCGGCGGCGGCGGCGG 13515

[illegible]

RESULT 12

US-08-458-076A-6
; Sequence 6, Application US/08458076A

; Patent No. 5698425

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Ulmer, Scott Joseph

APPLICANT: OKNER, SCOTT JOSEPH
TITLE OF INVENTION: Genes for the synthesis of

1 TITLE OF INVENTION: Genes for the synthesis of antipathogenic substances

TYPE OF INVENTION: 22
NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

; COUNTRY: USA

ZIP: 10532

; COMPUTER READABLE FORM:

```

;
; MEDIUM TYPE: Floppy disk

```

```
COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 110

```

APPLICATION NUMBER: US/08/458,076A

FILING DATE: 01-JUN-1999
CLASSIFICATION: 42E; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:

APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995

! FILING DATE: 01-JUN-1995
! APPLICATION NUMBER: 08/21

APPLICATION NUMBER: 08/238,261

; FILING DATE: 08-Jun-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Elmer, James Scott
 ; REGISTRATION NUMBER: 36,129
 ; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8614
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 28958 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-458-076A-6
 ;

Query Match

Best Local Similarity 60.5%; Pred. No. 2.3e-28;

Matches	323: Conservative	0: Mismatches	211: Indels	0: Gaps	0:
2000	2000	2000	2000	2000	2000

Qy	1	GGAA	CCAC	CCCT	CAT	CAC	CGG	GGAA	CCGG	CGC	CT	CGC	ACC	CA	CTT	CA	CCC	ACC	CA	CTC	60
Db	13456	GGG	AC	CGT	CT	CAT	CAC	GGG	AGCA	CCGG	GAC	GT	TAG	GAG	TCT	TGG	T	CG	CAC	GC	13515
Qy	61	ACC	AC	CC	CA	AA	CC	CA	CC	CA	CA	CT	CT	CT	CA	CC	AG	CG	AA	CG	120
Db	13516	GT	CG	GA	AA	CA	CAG	GG	CC	AA	CA	CT	GT	CT	CT	CA	CT	CG	AG	AG	13575
Qy	121	CAG	CA	AA	CA	CT	CA	CC	CA	CT	CA	AA	AA	AG	CA	TC	CA	CT	CA	CT	180
Db	13576	GG	CG	CG	AG	GC	T	CT	CG	AA	CG	AG	CT	CG	AA	GC	GT	GG	GG	GC	13635
Qy	181	TG	CG	AC	CAG	CA	AA	CC	CAG	CA	CT	CT	CA	AC	CA	CT	CT	CA	AC	CA	240
Db	13636	TG	CG	AG	TG	CC	CG	AA	CC	CG	GC	CT	CG	GA	CC	CT	CT	GG	AG	CA	13695
Qy	241	CG	CT	CA	CA	CG	T	CA	TC	CA	CC	AG	GA	T	CT	CG	AG	AG	CA	CA	300
Db	13696	CG	AT	CAC	GG	CG	T	CGT	CAC	CG	CG	CG	CC	CT	CG	AG	CA	GG	CG	CG	13755
Qy	301	AC	CC	CC	CA	CA	CT	CA	AA	GT	CT	CT	CG	CG	CA	AA	GC	CA	AG	CG	360
Db	13756	AG	GG	CG	AG	CG	AT	GT	CT	CG	GT	CT	T	T	GA	CC	CA	AG	CT	CG	13815
Qy	361	CA	A	T	C	A	C	C	C	C	T	C	A	C	G	C	T	T	G	T	420
Db	13816	GAG	CT	CAC	C	AG	G	AC	G	CG	GT	CG	CG	CT	T	CG	T	CT	T	CG	13875
Qy	421	T	T	CG	GG	CA	CC	CG	CA	AG	CA	AA	CT	A	G	CC	CG	CAG	CA	AG	480
Db	13876	C	T	T	G	T	G	T	C	A	G	T	C	A	G	T	C	A	G	T	13935
Qy	481	CAC	CA	CC	GG	CA	CC	CA	CA	CT	CC	CG	CA	CC	AG	CA	T	CG	CT	GG	534
Db	13936	CAT	C	CG	CG	CG	CC	AA	AG	ACT	CC	CA	CG	CT	T	CG	T	CG	CT	GG	13989

RESULT 13

US-08-764-233A-4

US-00-704-233A-4
; Sequence 4. Application US/08764233A

; Patent No. 5710

GENERAL INFORMATION:

APPLICANT: Ligon, James M.

APPLICANT: Schupp, Thomas

APPLICANT: Beck, James J.

APPLICANT: HILL, Dwight S.

APPLICANT: Neff, Snezana

APPLICANT: Ryals, John A.

; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

; NUMBER OF SEQUENCES:

Qy	181	TGCGACACGAGCAACCCAGACCAACTCCAAACAACTCCTCAAACACCATGCCCCCAACAAC	240
Db	12513	TGCGACGCGGCCAGACCCACGCGCCCTCCAAGCCCTTTGGACTCCATCCGACGCGTCAAC	12572
Qy	241	CCCTCACCAACCGTTCATCCACACGCGAGGCATCCTCGACGCGCACCCCTCACCAACCTC	300
Db	12573	CCGCTACCCGCGTGTCTACGCGCGGGGCCCTCGACGACGGCTGTCTGGCGCCATG	12632
Qy	301	ACCCCCACCAACTCAACACGTCCTCCGCGCCAAAGCCACAGCGGCCCACTCTCTCCAC	360
Db	12633	AGCCCCGAGCGCATCGACGCGTCTTTGCCCCAAAGCTCGATGCTGTTGGCACTTGCAT	12692
Qy	361	CAACTCACCAACACACACCCCCCTCAACGGCGTTGTCCTCTACTCTCGCGCGCGCCAC	420
Db	12693	GAGTCAACCAAGACAAGCCCTTCGCGCCTTGTCTCTTCTCGTCGCGTCTGGCGTC	12752
Qy	421	TTCGGCGCACCCGGCCAAAGCAACTACGCGGACGCAACGCGCTACTCTGACGCCCTCGCC	480
Db	12753	CTTTGTTAGTCCAGGTGAGTGAACACTACGCGGGCCATGCTTCTCTGATGCGCTCGG	12812
Qy	481	CACACGCGCACACCAACCACTTCCCGCCACACGATCGCTGGGGCACTGG	534
Db	12813	CATCACCGGCGTGCCACAGGGCTCCCGGCTCTCTCGCTCGCATGGGGTATTGG	12866

RESULT 7

```

US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

Query Match	36.8%;	Score 196.4;	DB 4;	Length 15872;
Best Local Similarity	60.5%;	Pred. No. 2.2e-28;		
Matches 323;	Conservative 0;	Mismatches 211;	Indels 0;	Gaps 0;
Qy	1	GGAAACACCCCTCATCAGCGGGGNAACGGGGCCCTCGCCACCCACCTCATCCACACCTC	60	
Db	7337	GGCACCCTCTCTGTCACCGGGGACCCGGAGGCTCGGAGCGTCTCTCGCCGCGCACCTC	7396	
Qy	61	ACACCCACCAACCCACCCCAACCTCTCTCTACAGCGCAACCGGCCGCCACACCCCC	120	
Db	7397	GTGAGCGCTCAGCGGGTGCGCCGCTCTCTGTGTGAGCCGCGCGGGGCGAGCGCCCC	7456	
Qy	121	CAGSCAAACACCTCACACCCCACTCCACAAAAGGCATCCACCTCACCATCACCAAC	180	
Db	7457	GGCGCGCGACCTGTGGCGAGGACCTCGCGGGCCCTCGGCGCGAGGTGGCGGCTTGGCGCC	7516	
Qy	181	TGGGACACCAAGCAACCCAGACCAACTCCAACTCTCTCAACACCATCCCGCCACACAC	240	
Db	7517	GCCGACCGCGGACCGCGAGAGCTGGCGGGCGGATCGCACCGGTGCGCGCGGAGCAT	7576	
Qy	241	CCCTCACCAACCGTTCATCCACACCGCAGGCGATCTCGACGACGCCACCTTCACCAACCTC	300	
Db	7577	CCGCTGACGGCGGTCTGTGCACACGGCGGGAGTCGTGACGACGGGACGTTGGAGGGGCTC	7636	
Qy	301	ACGCCCAACCAACTCAACAAAGTCCTCGCGCCAAAGCCACAGCGGCCACCTCTCTCCAC	360	

```

Db      7637  ACACCGGAAACGGCTGGAGCGGTACTTCGCGCCGAAAGGTGACACGCGCGGTGGAACCTTGCAC  7659
Qy      361  CAACCTACCAACAAACACACCCCTTACCGCCTTTCGTCTCTTACTTCCTCGCGCGCGCGCAACC  420
Db      7697  GAGCTCACCAAGAGACCTTCGCGCTTCGACGCGCTTCGTCTCTTCTCTCTCCGTCTCCGGCATC  7756
Qy      421  TTGGCGGCACCCCGGCCAAGCCAACTAGCGCGGAGCCAAACGCCTTACCTTCGACGCGCCTCGCC  480
Db      7757  GTCGGACACCGCGCGCCAGGCCAACTACGCGGCGGCCAAACACGGGCGCTTCGACGCGCCTCGCC  7816
Qy      481  CACCACCGCCACACCCACACCACTTCCCGCGCCACACAGCATCGCCTTGGGGGCACCTTGG  534
Db      7817  GCCACCGCGCGCCACCGCGGCGCTTGGCGCGCAACGTTCGTGCGCTGGGCGCTCTCTG  7870

RESULT 8
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

```

Query Match	36.8%	Score 196.4;	DB 1;	Length 28958;
Best Local Similarity	60.5%;	Pred. No. 2.3e-28;		
Matches 323; Conservative	0;	Mismatches 211;	Indels 0;	Gaps 0;

Qy	301	ACCCACACCCAACTCAACAAGTCTTCGGGGCAAAAGCCCAAGAGCCCACTCTCTCCAC	360
Db	18890	ACACCGCAGCGGCTCGCGGCGTCTCTCCGCCGAAGGCCGCGGTACGCAATCTGCAC	18949
Qy	361	CAACTCACCACACACACCCCTTCAACCGCTTCGTCTCTATCTCTCGCGCGCCACC	420
Db	18950	GAACCTACCCACAGGGGACGCGCTGTCTGGCGTTTCATCTCTATCTGTGCGCGCCGAGTG	19009
Qy	421	TTCTGGCGCACCCGGCCAAAGCAACTACGCGCAGCCCAACGCCCTACTCTGACGCCCTCGCC	480
Db	19010	CTGGCAGCGGGCCAGACGGGTACGCGCGCGCCACGCCCTACTTGGACTCTCTCGCC	19069
Qy	481	CACACCGCCACACCAACCACTCTCCCGGCAACAGCATCTGCTGGGGCACTGG	534
Db	19070	GTTGGCGGGAGCGGGGACTCGCCGCTATCTGCTCTGGCTGGGGCCCGTGG	19123

RESULT 6

US-08-764-233A-1
Sequence 1, Application US/08764233A
Patent No. 5716849
GENERAL INFORMATION:
APPLICANT: Ligon, James M.
APPLICANT: Schupp, Thomas
APPLICANT: Beck, James J.
APPLICANT: Hill, Dwight S.
APPLICANT: Neff, Snezana
APPLICANT: Ryals, John A.
TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-geigy Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,233A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/729,214
FILING DATE: 09-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: 1506/CIP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49177 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Sorangium cellulosum
IMMEDIATE SOURCE:
CLONE: p98/1, pJL3, and pVKM15
FEATURE:
NAME/KEY: misc_feature
LOCATION: 383..760

Db 6982 GCTCGCGCGCGGACACCGCGCGGTCCGCGCTCTCGGTGGGACCGGTGG 7035

RESULT 4
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-34

Query Match 40.1%; Score 214; DB 4; Length 4689;
Best Local Similarity 63.9%; Pred. No. 1.2e-31;
Matches 341; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

Qy 1 GGAAACCAACCTCATCAGCGGCGGAACCGGCGGCTCTCGCCACCCACCTTCCACCACCACTC 60
Db 3346 GGCACCGTCTCATCAGCGGCGGACCGGAGCCCTCGGCGACCAAGCGCCACGCTGGATG 3405
Qy 61 ACCACCCACCAACCCACCCCAACACTCTCTCTCTACACGAGCGGAACCGGCGCCCGCCACACCCC 120
Db 3406 GCCACCAACGAGGCGG---AACACTCTCTCTCTAGCGCGAGCGGCGGAACAGCCCCC 3462
Qy 121 CAGCACAACACACTCACCACCCCAACTCCAAACAAAGGCATCCACCTTCACCATCACCACACC 180
Db 3463 GGAGCCACCCCAACTCAGCGCGAACTCAGCGCATCGGCGCGCCGCGTCAACATCGCGCC 3522
Qy 181 TGGACACCAAGCAACCCAGAGCAACTCCAACTCTCTAACACCATCTCCCGCCCGCCACACAC 240
Db 3523 TGGACATCGCCGACCCCGCCAGCCGCAATCGCACCTCTCTGACGCGCATCCCGCGCGAGCG 3582
Qy 241 CCCCTCACCACCGGTATCCACACGCGAGGATCTCTCGACGAGCCACCTCACCACCACTC 300
Db 3583 CCCCTCAGCGGTGTCTCAACCGCGCGGCAACCGCGCGGCGATCGCTTGAAGCTCACC 3642
Qy 301 ACCCCACCCCAACTCAACAAAGTCTCTCGCGGCCAAAGCCACAGCGGCCACCTCTCTCCAC 360
Db 3643 GGCGCGAGGACATCGCCGATCTCTGGCGCGAAGACGAGCGGCGCGAGGTCTCTGAC 3702
Qy 361 CAATCAACCAACACACCCCGCTTCAACCGCTTCTGTCTCTTACTCTCTCGCGCGCGCAACC 420
Db 3703 GACCTGTCTCGCGGCACTCGCGCTGGAGCGCTTGTCTCTTACTCTCTCGAACGCGGGGTC 3762
Qy 421 TTGGCGCACCCGGCGCAAGCAACTAGCGCGAGCCAAACGCTTACTCTGAGCGCTCTCGCC 480
Db 3763 TGGGGCAGCGGACGAGGCGGTCTAGCGGCGGCGCAACGCCCACTCTGACGCGCTCGCC 3822
Qy 481 CAGCACCGCCACACCCACCACTCTCCCGGCCACACAGCATCGCTTGGGGCACTTGG 534
Db 3823 GCCCGGCGCGCGCGGCGGAGCGGACCTCGGTCTGCTTGGGCTCTTGG 3876

RESULT 5
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.

Db 6979 GAGCTGACCCGGGACCTCGACCTCGAGCGGTTCTGCTCTTCTCGTCCGTGCGAGCACT 7038
Qy 421 TTGCGGCAACCGGCAAGCCAACTAGCGCGGAGCAAGCCCTACTCGAGCCCTCGCC 480
Db 7039 CTGGGATCCCGGTGAGGCAACTAGCGCCGCGCAACAGCCCTACTCGAGCCCTCGCG 7098
Qy 481 CACCAACCGGCAACCCACCACTCCCGCGCACAGCATCGCTGGGGACCTGG 534
Db 7099 GCTGCGCGCGGCGCACCGCGCGTCCCGCGTCTCGTGGCTTGGGACCGTGG 7152

RESULT 2

US-09-105-537-5

; Sequence 5, Application US/09105537A

; Patent No. 6285202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 5

; LENGTH: 36778

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-105-537-5

Query Match 41.9%; Score 223.6; DB 4; Length 36778;

Best Local Similarity 65.0%; Pred. No. 2.4e-33;

Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 1 GGAAACCACTCTATACCGCGGGAACCGCGCCTCGCCACCCACCTGACCCACCACTC 60
Db 8363 GGCAACGCTCTCTATACCGCGGGAACCGCGCCTCGCCACCCACCTGACCCACCACTG 8422
Qy 61 ACCACCAACCAACCAACCAACCACTCTCTCTCACAGCGGAACCGCGCCCAACACCC 120
Db 8423 GCCACCAACCGGAGCGG---AACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8479
Qy 121 CAGGCAACACCTCTACCACT 180
Db 8480 GGAGCCACCACTCTACCACT 8539
Qy 181 TGGGACACGAGCAACCAACCACT 240
Db 8540 TGGGAGCTGCGGACACCAACCACT 8599
Qy 241 CCCTTACCACT 300
Db 8600 CCCTTACCACT 8659
Qy 301 ACCCCCAACCACT 360
Db 8660 ACCCGGAGAGGTCTCGCGGCGCCACCGTGTGGAAGGCGCGTCTCTCTCTCTCTCTCTCTCT 8719
Qy 361 CAACTACCAACCACT 420
Db 8720 GAGTACCGCGGACCT 8779
Qy 421 TTGCGGCAACCGGCAAGCAACTAGCGCGGAGCAAGCCCTACTCTGAGCCCTCTCGCC 480
Db 8780 CTGGGATCCCGGTGAGGCAACTAGCGCCGCGCAACAGCCCTACTCTGAGCCCTCTCGCG 8839
Qy 481 CACCAACCGGCAACCACT 534
Db 8840 GCTCGCGCGGCGCAACCGCGCGTCT 8893

RESULT 3

US-09-320-878-19

; Sequence 19, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Melanie C.

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MEDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 19

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-320-878-19

Query Match 41.9%; Score 223.6; DB 3; Length 38506;

Best Local Similarity 65.0%; Pred. No. 2.4e-33;

Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 1 GGAAACCACTCTATACCGCGGGAACCGCGCCTCGCCACCCACCTGACCCACCACTC 60
Db 6505 GGCAACGCTCTCTATACCGCGGGAACCGCGCCTCGCCACCCACCTGACCCACCACTG 6564
Qy 61 ACCACCAACCAACCAACCAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 6565 GCCACCAACCGGAGCGG---AACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6621
Qy 121 CAGGCAACCACTCTACCACT 180
Db 6622 GGAGCCACCACTCTACCACT 6681
Qy 181 TGGGACACGAGCAACCAACCACT 240
Db 6682 TGGGAGCTGCGGACACCAACCACT 6741
Qy 241 CCCTTACCACT 300
Db 6742 CCCTTACCACT 6801
Qy 301 ACCCCCAACCACT 360
Db 6802 ACCCGGAGAGGTCTCGCGGCGCCACCGTGTGGAAGGCGCGTCTCTCTCTCTCTCTCTCTCT 6861
Qy 361 CAACTACCAACCACT 420
Db 6862 GAGTACCGCGGACCT 6921
Qy 421 TTGCGGCAACCGGCAAGCAACTAGCGCGGAGCAAGCCCTACTCTGAGCCCTCTCGCC 480
Db 6922 CTGGGATCCCGGTGAGGCAACTAGCGCCGCGCAACAGCCCTACTCTGAGCCCTCTCGCG 6981
Qy 481 CACCAACCGGCAACCACT 534

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 23.1323 Seconds
(without alignments)

7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676

Perfect score: 534

Sequence: 1 ggaacacacctcatcacgg.....gcacgctggggcacctgg 534

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/prodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/prodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/prodata/1/ina/PCUTUS_COMB.seq.*

6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	223.6	41.9	13842	4	US-09-105-537-30
2	223.6	41.9	36778	4	US-09-105-537-5
3	223.6	41.9	38506	3	US-09-320-878-19
4	214	40.1	4689	4	US-09-105-537-34
5	210.8	39.5	43280	2	US-08-804-227C-1
6	207.6	38.9	49377	1	US-08-764-233A-1
7	196.4	36.8	15872	4	US-09-105-537-1
8	196.4	36.8	28958	1	US-08-258-261B-6
9	196.4	36.8	28958	1	US-08-456-837-6
10	196.4	36.8	28958	1	US-08-457-342-6
11	196.4	36.8	28958	1	US-08-457-646A-6
12	196.4	36.8	28958	1	US-08-458-076A-6
13	196.4	36.8	28958	1	US-08-764-233A-6
14	196.4	36.8	28958	1	US-08-457-335A-6
15	196.4	36.8	28958	1	US-08-729-214-6
16	196.4	36.8	28958	3	US-09-028-934-6
17	191.6	35.9	44377	2	US-08-804-227C-7
18	191.6	35.9	44377	2	US-08-804-198-1
19	185.2	34.7	11220	4	US-09-105-537-32
20	180.4	33.8	13987	2	US-08-804-227C-13
21	177.2	33.2	11219	1	US-07-642-734C-1
22	177.2	33.2	11219	3	US-08-439-009A-1
23	175.2	32.8	477	3	US-09-010-809-2
24	175.2	32.8	477	3	US-09-010-809-17
25	154.6	29.0	50937	4	US-09-428-517-1
26	129.2	24.2	80161	3	US-09-036-987A-1
27	129.2	24.2	80161	4	US-09-370-700-1

28	128	24.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
29	128	24.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
30	124.2	23.3	33529	4	US-09-144-085-3	Sequence 3, Appli
31	119.2	22.3	925	3	US-08-858-003-1	Sequence 1, Appli
32	119.2	22.3	925	3	US-09-078-166-1	Sequence 1, Appli
33	119.2	22.3	925	4	US-08-997-467-1	Sequence 1, Appli
34	117.4	22.0	897	4	US-09-434-288-6	Sequence 6, Appli
35	117.4	22.0	1926	4	US-09-249-585A-4	Sequence 4, Appli
36	117.4	22.0	1931	2	US-09-130-114-2	Sequence 2, Appli
37	110.4	20.7	68750	3	US-09-335-409-1	Sequence 1, Appli
38	110.4	20.7	68750	4	US-09-568-102-1	Sequence 1, Appli
39	110.4	20.7	68750	4	US-09-567-969-1	Sequence 1, Appli
40	110.4	20.7	68750	4	US-09-568-480-1	Sequence 1, Appli
41	110.4	20.7	68750	4	US-09-568-486-1	Sequence 1, Appli
42	110.4	20.7	68750	4	US-09-568-472-1	Sequence 1, Appli
43	110.4	20.7	68750	4	US-09-567-899-1	Sequence 1, Appli
44	110.4	20.7	71989	4	US-09-443-501A-2	Sequence 2, Appli
45	110	20.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-105-537-30

; Sequence 30, Application US/09105537A

; Patent No. 6265202

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 30

; LENGTH: 13842

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-105-537-30

Query Match	41.9%;	Score	223.6;	DB	4;	Length	13842;		
Best Local Similarity	65.0%;	Pred. No.	2.3e-33;						
Matches	347;	Conservative	0;	Mismatches	184;	Indels	3;	Gaps	1;
QY	1	GGAAACACCTCATACCGGGGGAACCGGCGCCCTCGCCACCCACCTCACCCACCAACCTTC	60						
Db	6622	GGCAGCGTCTCATCACCGGGGGAACCGGAGCCCTCGGCAGCCACCGCGCACGCTGGATG	6681						
QY	61	ACCAACCAACCAACCAACCAACCTCTCTCTACAGCGGAACGGGGCCCCACACCCCC	120						
Db	6682	GCCACCAACGGAGCGG--AACACCTCTCTCTCTGTCAGCGGAGGGGCGAACAAGCCCC	6738						
QY	121	CAGGACAACACTCACCAACCCAACTCCAAACAAAAGGCATCCACTACCATCACCACC	180						
Db	6739	GGAGCAACCAACTCACCGCCGAACTCACCGATCGGGCGCGCGGTACCATCGCGGCC	6798						
QY	181	TGCGACACCAAGCAACCCAGACCAACTCCAAACACTCTCTCAACACCATCCCCCCCAACAC	240						
Db	6799	TGCGAGTGCAGACCCCAACCCGCAATGGGCACCTCTCGAGCGCCATCCCCGCCGAGACG	6858						
QY	241	CCCTTACCAACCGTATCACACCGCAGGACATCTCTCGACGAGCGCACCTTCACCAACCTC	300						
Db	6859	CCCTTACCGCGGTGTCTCACACCGCGCGGCGCTCGACAGCGGCATCTGTGACACGCTG	6918						
QY	301	ACCCCAACCAACTCAACAAACGTCTCGCGCGCAAAAGCCCAACAGCGGCCACCTCTCTCCAC	360						
Db	6919	ACCGGAGCAGGTTCGGGGGGCCACACGTGCGAAGGCCGTTCGGCGCCTCGGTGCTCGAC	6978						
QY	361	CAACTCACCAACACACACCCCCCTTACCGCTTTCGTCTCTACTCTCTCGCGCGCGCCACC	420						

THIS PAGE BLANK (USPTO)

CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.

XX

SQ Sequence 4571 BP; 659 A; 1829 C; 1469 G; 614 T; 0 other;

Query Match 45.7%; Score 244; DB 21; Length 4571;
Best Local Similarity 67.7%; Pred. No. 3.1e-31;
Matches 359; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

QY	5	CGACCTCTCATCCGCGGGAACCGCGGCTCGCCACCCACCTCACCACCCACCTCACCACCA	64
Db	3943	CCCTCCACACCAACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCA	4002
QY	65	CCACCAACACCCACCCACCCACCTCTCTCACCAGCCGAACCGGCGCCACACCCCCACG	124
Db	4003	TCACCGGCGGCTCGGACCCCTCGCGGCATCTCTCGCGGCACCTGACACCCCCACA	4062
QY	125	CACAACACCTCACCACCCAACTCCAAACAAAGGCAATCCACTCACCATCACCACCTGG	184
Db	4063	CCTACCTCTCTCCGCAACCCACCCCGGACCGCCACCCCGGCAACCCACCTCCCTGG	4122
QY	185	ACACGACGACCCAGACCACTCCAACTCTCTGACGACATCCCGCCACACACCCCC	244
Db	4123	AGTGGGCGACCCCGCCACCACTGGCCACCACTCACCACATCCCGCCCAACCC	4176
QY	245	TCACCACTCTCATCCACACGCGGATCTCTGACGACGCGCCCTCACCACCTCACCC	304
Db	4177	TCACCGGCATCTTCCACACGCGGCGCCCTCTGACGACGCGCATCTCCAGCGCTCACCC	4236
QY	305	CCACCCAACTCAACAAAGTCTCTCGCGCCAAAGCCACAGCGCCCACTCTCTCCACCAAC	364
Db	4237	CGACCGCTCACCACCGTCTCTCCACCCCAAGCCAAACGCGCTGGCACTTGACCAAC	4296
QY	365	TCACCCAAACACACCCCGCTCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	424
Db	4297	TCACCCAAACCAACCCCTCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	4356
QY	425	GGCACCCCGGCAAGCAACTAGCGCGAGCAAGCCCTACCTCGACGCGCTCGCCACC	484
Db	4357	GCAGCCCGGACAGGAATACTAGCGCGCGCCCAAGCCCTCTCTGACGCGCTCGCCACC	4416
QY	485	ACGCGCACACCAACCACTCCCGGCCACCAAGCATCGCCTGGGGCACCTGG	534
Db	4417	ACGCGCACACCTCGGCGCAACCGCCACCTCCATCGCCTGGGGCATGTGG	4466

Search completed: June 17, 2003, 13:48:28
Job time : 109.318 secs

DT	08-AUG-2000	(first entry)
XX	Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.	
XX		
XX	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant; immunophilin; FK-506 binding protein; polyketide compound; uveitis; transplant rejection; graft-versus-host disease; alopecia universalis; autoimmune chronic active hepatitis; inflammatory bowel disease; multiple sclerosis; primary biliary cirrhosis; scleroderma; neurite outgrowth; nerve regrowth; Parkinson's disease; Alzheimer's disease; stroke; traumatic spinal cord, brain injury; peripheral neuropathy; ss.	
OS	Synthetic.	
OS	Streptomyces hygroscopicus.	
XX		
XX	Key	Location/Qualifiers
XX	CDS	3..4466
XX		/*tag= a
XX		/note= "no termination codon given"
XX	WO200020601-A2.	
XX		
XX	13-APR-2000.	
XX		
XX	01-OCT-1999;	99WO-US22886.
XX		
XX	02-OCT-1998;	98US-0102748.
XX	11-MAR-1999;	99US-0123810.
XX	17-JUN-1999;	99US-0139650.
XX	(KOSA-) KOSAN BIOSCIENCES INC.	
XX		
XX	Reeves C, Chu D, Khosla C, Santi D, Wu K;	
XX		
XX	WPT; 2000-317716/27.	
XX	P-PSDB; RAY84725.	
XX	New isolated polyketide synthase nucleic acid and polyketide compounds, useful for treating e.g. transplant rejection, uveitis, multiple sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or peripheral neuropathy -	
XX	Example 1; Page 76-79; 126pp; English.	
XX		
XX	The present sequence represents an AvrII-XhoII fragment that encodes module 8 of the FK-520 polyketide synthase (PKS) gene cluster with the endogenous acyltransferase (AT) domain replaced by the AT domain of module 12 of the rapamycin PKS (which is specific for malonyl units). FK-506 is a potent immunosuppressant, and acts through initial formation of an intermediate complex with protein immunophilins known as FK-506 binding proteins. The nucleic acids are used for producing polyketide compounds. The polyketide compounds can be used as immunosuppressants to prevent or treat transplant rejection, graft-versus-host disease or uveitis. They can also be used for treating e.g. alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, or scleroderma. They also have neurotrophic activity and can be used to promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. They can also be used for treating e.g. Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or peripheral neuropathies. They can also be used in agricultural and veterinary applications.	
XX		
XX	Sequence 4478 BP; 632 A; 1757 C; 1465 G; 624 T; 0 other;	
XX		
XX	Query Match	45.7%; Score 244; DB 21; Length 4478;
XX	Best Local Similarity	67.7%; Pred.No. 3.1e-31;
XX	Matches 359; Conservative	0; Mismatches 165; Indels. 6; Gaps 1;
QY	5 CCACCCCTCATCACGGCGGGAACGGCGCCCTCGCCACCCACCTCACCCACCACTCACCA 64	

CC compounds can be used as immunosuppressants to prevent or treat
CC transplant rejection, graft-versus-host disease or uveitis. They can
CC also be used for treating e.g. alopecia universalis, autoimmune
CC chronic active hepatitis, inflammatory bowel disease, multiple
CC sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
xx
SQ Sequence 4767 BP; 731 A; 1945 C; 1468 G; 623 T; 0 other;

Query Match	48.2%;	Score 257.4;	DB 21;	Length 4767;								
Best Local Similarity	67.8%;	Pred. No. 2.1e-33;										
Matches 360;	Conservative 0;	Mismatches 171;	Indels 0;	Gaps 0;								
Qy	4	ACCA	CCCTCATC	ACGGGGAAACGGGCGCCCTCGGCCACCCACTCTCA	CCCA	CCCACTCACC	63					
Db	4140	ACCA	CCCAACAACACCA	CAACCAACCCCAACACCCCAACCCCTTCA	CCCCCA	CCCAACCA	4199					
Qy	64	ACCA	CCAAACCC	ACCAACCTCTCTCTCA	CCAGCGGAAACGGCGCCCA	CCCAACCCCCCAC	123					
Db	4200	ATCT	CATCAC	CGGGGCTCGGCACCTTC	CGCGGCATCTCGCGGCACCTCA	CCAC	4259					
Qy	124	GC	ACAACCTT	CAACCAACCACTCA	CAAAAGGATTC	CACTCA	GCATCA	CACTGC	183			
Db	4360	CCCC	ACACTT	ACTCTCTCCG	CACACACCAACCC	CCCCCA	CCACACCGGC	CCCACTC	4319			
Qy	184	GAC	ACGAC	CCAGCA	CACTCC	CAACCTCTCA	ACCACTTC	CCCCCA	CAACACCC	243		
Db	4320	CCCT	GGACCTT	CACCGACCCCA	CCCAATCA	CCCAAGCCTT	CA	CCCCACAT	TACCA	4379		
Qy	244	CTCA	CCACCGT	CATCTCA	CACCGCAGGATCT	TCGACGACGCA	CCCTCA	CCCAACCT	CACC	303		
Db	4380	CTCA	CCGGCATCTT	CCACACCGCGCC	ACCTTC	CGACGAGCC	ACCTTCA	CCCACT	CACC	4439		
Qy	304	CCCA	CCCACTCA	ACAGTCTCTCG	GGCGCAAGGCC	CAACGCGCC	CACTCT	CTCA	CCAA	363		
Db	4440	CCCC	AAACCTT	CACCAACACCTT	CAACCCCAACCGCGCT	GGCACTT	CCAC	CCAC		4499		
Qy	364	CTCA	CCCAACAC	ACCCCTTCA	CGCTTCGTCTT	ACTCTTC	CGCGGGC	CACTTC		423		
Db	4500	CACA	CCCAAA	ACCAACCCCTT	CACCACTT	TGCTCTTACT	CCAGCGCGG	CACTTC		4559		
Qy	424	GGCG	CAACCGG	CAAGCAACT	ACGCGCGAC	CAACGCTT	ACTTC	GACG	CCCTCG	CCAC	483	
Db	4560	GGCA	CGCCG	CAAGCC	CACTA	CGCGCGC	CAAGCCTT	CTCT	CGAG	CGCTCG	4619	
Qy	484	CAC	CGCA	CA	CCCA	CACTTCC	CGCCCA	CCAG	CA	TCCCTGGG	CACTCG	534
Db	4620	CAC	CGC	CA	CA	CCCA	CA	CCCA	CA	CACTCG	CGCTGGG	4670

RESULT 11	
AAAL14669	ID AAA14669 standard; DNA; 4818 BP.
XX	
XX	
AC	AAA14669;
XX	
DT	08 -AUG- 2000 (first entry)
XX	
DE	Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
XX	
KW	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW	immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW	transplant rejection; graft-versus-host disease; alopecia universalis;
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW	neurile outgrowth; nerve regrowth; Parkinson's disease;

Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
peripheral neuropathy; ss.

Synthetic.

Streptomyces hygroscopicus.

Key Location/Qualifiers
CDS 3..4817
/*tag= a
/note= "no termination codon given"

W0200020601-A2.

13-APR-2000.

01-OCT-1999; 99WO-US22886.

02-OCT-1998; 98US-0102748.
11-MAR-1999; 99US-0123810.
17-JUN-1999; 99US-0139650.

(KOSA-) KOSAN BIOSCIENCES INC.

Reeves C, Chu D, Khosla C, Santi D, Wu K;
WPI; 2000-317716/27.
P-PSDB; AAY84733.

New isolated polyketide synthase nucleic acid and polyketide compounds,
useful for treating e.g. transplant rejection, uveitis, multiple
sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
peripheral neuropathy

Example 2; Page 102-105; 126pp; English.

The present sequence represents module 8 of the FK-520 polyketide
synthase (PKS) gene cluster, containing the acyltransferase (AT)
domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
and acts through initial formation of an intermediate complex with
protein immunophilins known as FK-506 binding proteins. The nucleic
acids are used for producing polyketide compounds. The polyketide
compounds can be used as immunosuppressants to prevent or treat
transplant rejection, graft-versus-host disease or uveitis. They can
also be used for treating e.g. alopecia universalis, autoimmune
chronic active hepatitis, inflammatory bowel disease, multiple
sclerosis, primary biliary cirrhosis, or scleroderma. They
also have neurotrophic activity and can be used to promote neurite
outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
and in intact animals, they promote regrowth of damaged facial and
sciatic nerves, and repair lesioned serotonin and dopamine neurons in
the brain. They can also be used for treating e.g. Parkinson's disease,
Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
peripheral neuropathies. They can also be used in agricultural and
veterinary applications.

Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 Other;

	Query Match	48.2%;	Score 257.4;	DB 21;	Length 4818;
	Best Local Similarity	67.8%;	Pred. No. 2.1e-33;		
	Matches 360;	Conservative 0;	Mismatches 171;	Indels 0;	Gaps 0;
QY	4	ACCACCTCATCACGGCGGAAACGGGGCGCTCGCCACCCACCTCTACCCACCACTCTACC	63		
Db	4191	ACCACCAACAACACCAACCAACACCCCTCCAAACACCCCAACCCCTCAACCCCAACCAAGCC	4250		
QY	64	ACCCACCAACCCCAACCAACACTCTCTCTTCAACAGCGGAAACGGCGCCCAACACCCGCCAC	123		
Db	4251	ATCTCTATACCGGGGGCTTCGGACCTCTGCGGGATCTCTGCGCGCACTCTCAACCAAC	4310		
QY	124	GCACAAACACTCACCAACCAATCTCAACAAAAAGGATCCACTCAACCACTGC	183		
Db	4311	CGCCACCACTACTCTCTCTCCGGCAACCAACACCCCAACCAACACCGGCAACCAATC	4370		


```
XX SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 50.0%; Score 266.8; DB 22; Length 125401;
Best Local Similarity 68.7%; Pred. No. 5.4e-35;
Matches 367; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 1 GGAACCAACCTATCACCGCGGGAACCGCGCGCTCGCCACCCACCTCAACCCACCACTTC 60
Db 40009 GGCACCGTGTCTCTACCGCGGCGACCGGGGCTGTGGCGGCTGTCTGCGCGCACTG 40068

QY 61 ACACCCACCAACCCACCAACCTCTCTCTCACCAGCGGAACCGCGCGCCACACCCCC 120
Db 40069 GTACCGGAGCACCGCGCGCGCCACCTGCTGTGGCGCGCGCGCGCGCGCGCGCGCC 40128

QY 121 CAGCGACACACCTCACCAACCCCACTCTCCACAAAGGATCTCACTCTCACTACCACTC 180
Db 40129 GGGCGCGCGGCACTCCACGCGCGAACTGACCGCGCTGGGGCGCGAGTCACTGCGCGC 40188

QY 181 TGGGACACGACGACCAACCCAGACCACTCTCAACAACTCTCAACACCACTCCCGCCACACAC 240
Db 40189 TGGGAGTTCGCGACCGCGCGGCTGTGGCGGCTGTGTGCGACCGTGTGGCGCGGACAC 40248

QY 241 CCCTCACCACTATCCACACCGGAGGATCTCGACGAGCCACCTCGACGAGCCACCTTCACCACTTC 300
Db 40249 CCCTCACCGGCTGTGTGACACCGCGCGGTCTGTGACGAGCGGACCTTCACCGGCTG 40308

QY 301 ACCCCACCACTCAACAGCTCTCCGGCGCGCAAGCCACAGCGCCACCTCTCTCCAC 360
Db 40309 AACCCGACCGCTCGCCACCGCTCTACGCGCCAGGTGGACGCGCCCTGGCACCTGAC 40368

QY 361 CAACTCACCAACACACACCCCGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 40369 GACCTCACCGGCACTTCGACCTGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 40428

QY 421 TTGGGCGCACCGGCGCAAGCACTACGCGCGGCAACGCTTACCTCGACGCGCTTCGCC 480
Db 40429 ATGGGCGGACCGGCGGCGGCACTACGCGCGGCAACACCTTCTCTCGACGCGCTTCGCC 40488

QY 481 CACCGGCGCACACCCACCTTCGCGCGGCAACGATCGCTGGGGGCACTGG 534
Db 40489 GCCACCGACACGCGCTCGCGCTGTGGCGGCACTCTGCTGGCGGCGCGCTGG 40542

RESULT 6
AD17184
ID AAD17184 standard; DNA; 65140 BP.
AC AAD17184;
XX
XX
XX
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nysl DNA of nystatin PKS gene cluster.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nysl; ds.
XX
XX Streptomyces noursei.
XX
XX Key Location/Qualifiers
XX CDS complement (1..1035)
XX FT /tag= a
XX FT /product= "NysD2 partial protein"
XX FT /note= "CDS does not include stop codon"
XX FT complement (1056..2576)
XX FT /tag= b
XX FT /product= "NysD1 protein"
XX FT 2806..6906
XX FT /tag= c
XX FT /product= "NysA protein"
XX FT 6952..16530
XX FT /tag= d
XX FT /product= "NysB protein"
```

```
FT CDS 16550..49840
FT FT /tag= e
FT FT /product= "NysC protein"
FT CDS 50260..51015
FT FT /tag= f
FT FT /product= "NysE protein"
FT CDS 51405..54305
FT FT /tag= g
FT FT /product= "NysR1 protein"
FT CDS 54329..57190
FT FT /tag= h
FT FT /product= "NysR2 protein"
FT FT /note= "CDS does not include start codon"
FT CDS 57180..59963
FT FT /tag= i
FT FT /product= "NysR3 protein"
FT FT 60415..61047
FT FT /tag= j
FT FT /product= "NysR4 (short) protein"
FT FT /note= "CDS does not include start codon"
FT CDS 61736..62497
FT FT /tag= k
FT FT /product= "NysR5 protein"
FT FT /note= "CDS does not include start codon"
FT CDS complement (62551..63615)
FT FT /tag= l
FT FT /product= "ORF2 protein"
FT FT /note= "CDS does not include start codon"
FT CDS 63765..64961
FT FT /tag= m
FT FT /product= "ORF1 protein"
FT FT WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
XX WPI: 2001-557614/62.
XX P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
XX AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nysl DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
```

XX	Mutated Type I polyketide synthase containing a polylinker site in an
PT	extension module for replacement of a reductive loop sequence, for
PT	producing polyketides, e.g. B1 avermectin -
XX	
PS	Disclosure; Fig 7a-f; 75pp; English.
XX	
CC	The present sequence is that of DNA encoding the first 2 modules
CC	of the avermectin polyketide synthase (PKS) of Streptomyces
CC	avermitilis. The invention relates to nucleic acids encoding a
CC	Type I PKS such as avermectin in which a polylinker with multiple
CC	restriction sites replaces or 1 more PKS genes encoding enzymes
CC	associated with reduction. Novel PKS are provided in which in
CC	which the reductive loop in a selected module of the Type I PKS is
CC	replaced with the equivalent segment from the same or different
CC	PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC	host cells, and methods for producing novel polyketides by
CC	culturing host cells are claimed. The polyketides obtained are
CC	useful as antibiotics and insecticides. Fermentation products
CC	containing C22-C23 dihydroavermectin, ivermectin and B1
CC	avermectins are claimed.
XX	
SQ	Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
	Query Match 99.7%; Score 532.4; DB 21; Length 12381;
	Best Local Similarity 99.8%; Pred. No. 9.8e-76;
	Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 GGAAACCACTTCATCACCGGCGGAACCGGCGCCTTCTCTCAACAGCGAACC
DB	5960 GGAACCACTTCATCACCGGCGGAACCGGCGCCTTCTCTCAACAGCGAACC
	6020 ACCACCCACCAACCCACCCCAACACTCTCTCTCACAGCGGAACCGGCGC
QY	61 ACCACCCACCAACCCACCCCAACACTCTCTCTCACAGCGGAACCGGCGC
DB	6020 ACCACCCACCAACCCACCCCAACACTCTCTCTCACAGCGGAACCGGCGC
QY	121 CAGGCACAACACTCACCACCCCAACTCCACAAAAGCATCGACATCAGATCACC
DB	6080 CAGGCACAACACTCACCACCCCAACTCCACAAAAGGCATCGACATCAGATC
QY	181 TGGGACACAGCAACCCAGACCAACTCCAACAACTCTCTCAACACCATCCCC
DB	6140 TGGGACACAGCAACCCAGACCAACTCCAACAACTCTCTCAACACCATCCCC
QY	241 CCCTCTACCAACCGTGATCCACACCGCAGGCATCTTCGACGACGCCCTTCA
DB	6200 CCCTCTACCAACCGTGATCCACACCGCAGGCATCTTCGACGACGCCCTTCA
QY	301 ACCCCGACCAACTCAACAACGTCTTCGGCGCAAAGCCACAGGCGCATCTCT
DB	6260 ACCCCGACCAACTCAACAACGTCTTCGGCGCAAAGCCCAAGCGGCCCATCT
QY	361 CAACTCACCCAACACACCCCCTTCCGCGCTTCTACTCTCTCGCGCGCGCACC
DB	6320 CAACTCACCCAACACACCCCCTTCCGCGCTTCTACTCTCTCGCGCGCGCACC
QY	421 TTCCGCGCATCCCGSCCAAGCCAACTACGCGCGCAGCCAAACGCTACTCTG
DB	6380 TTCCGCGCATCCCGSCCAAGCCAACTACGCGCGCAGCCAAACGCTACTCTG
QY	481 CACCACCGCCACACCCACCACTCTCCCGCCACACAGCATCGCTGGGGCACT
DB	6440 CACCACCGCCACACCCACCACTCTCCCGCCACACAGCATCGCTGGGGCACT

DE	Streptomyces noursei nystatin PKS gene cluster DNA.
XX	
KW	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW	antifungal; antibiotic; ds.
OS	Streptomyces noursei.
XX	
Key	Location/Qualifiers
FT	6337..34771
FT	/tag= a
FT	/product= "NysI complete protein"
FT	34792..51099
FT	/tag= b
FT	/product= "NysJ protein"
FT	51155..57355
FT	/tag= c
FT	/product= "NysK protein"
FT	57503..58687
FT	/tag= d
FT	/product= "NysL protein"
FT	complement (58786..58980)
FT	/tag= e
FT	/product= "NysM protein"
FT	/note= "CDS does not include start codon"
FT	complement (59045..60241)
FT	/tag= f
FT	/product= "NysN protein"
FT	/note= "CDS does not include start codon"
FT	complement (60238..61296)
FT	/tag= g
FT	/product= "NysO2 complete protein"
FT	120628..121308
FT	/tag= h
FT	/product= "NysR4 (long) protein"
XX	
PN	WO200159126-A2.
XX	
PD	16-AUG-2001.
XX	
PP	08-FEB-2001; 2001WO-GB00509.
XX	
PR	08-FEB-2000; 2000GB-0002840.
PR	10-APR-2000; 2000GB-0008786.
PR	14-APR-2000; 2000GB-0009387.
XX	
PA	(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA	(SNTF) SINTEP STIFTELSEN IND TEK FORSK.
PA	(ALPH-) ALPHARMA AS.
PA	(SINV-) SINVENT AS.
PA	(DZIE/) DZIEGLEWSKA H.
PA	(ZOTC/) ZOTCHEV S B.
PA	(SEKU/) SEKUROVA O N.
PA	(FJAE/) FJAERVIK E.
PA	(BRAU/) BRAUTASET T.
PA	(STRO/) STROM A R.
XX	
PI	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI	Valla S, Ellingren TE, Sletta H, Gulliksen O;
XX	
DR	WPI; 2001-557614/62.
DR	P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148;
DR	AAE10149, AAE10150.
XX	
XX	New nystatin polyketide synthase polynucleotides and polypeptides,
PT	useful as antibiotics and antifungals -
XX	
PS	Claim 1; Page 188-254; 266pp; English.
XX	
CC	The present invention relates to the cloning and sequencing of the
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC	involved in the biosynthesis of the macrolide antibiotic nystatin.
CC	The nystatin PKS is useful as antifungal antibiotics. The present
CC	sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.

QY 301 ACCCCACCACTCAACAGCTCTCCGCGCAAGCCACAGCGCCACCTCTCCAC 360
Db 5443 ACCCCACCACTCAACAGCTCTCCGCGCAAGCCACAGCGCCACCTCTCCAC 5502
QY 361 CAACTCACCACACACCCCTTACCGGCTTCTCTACTCTCTCCGCGCGGCACC 420
Db 5503 CAACTCACCACACACCCCTTACCGGCTTCTCTACTCTCTCCGCGCGGCACC 5562
QY 421 TTCCGGCGCACCCGCGCAAGCAACTAGCGCGCAGCAAGCCCTACCTCGAGCGCTCGCC 480
Db 5563 TTCCGGCGCACCCGCGCAAGCAACTAGCGCGCAGCAAGCCCTACCTCGAGCGCTCGCC 5622
QY 481 CACCACCGGCACACCCACCACTCTCCGCGCACAGCATCGCTTGGGCGACCTGG 534
Db 5623 CACCACCGGCACACCCACCACTCTCCGCGCACAGCATCGCTTGGGCGACCTGG 5676

RESULT 3

AAH79277
ID AAH79277 standard; DNA; 30690 BP.

AC AAH79277;

DT 04-DEC-2001 (first entry)

DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.

XX Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; ds.

OS Streptomyces avermitilis.

XX Key Location/Qualifiers
FH CDS 1..11919
FT /*tag= a
FT /product= "AAG65264"
FT /partial
FT CDS 11971..30690
FT /*tag= b
FT /product= "AAG65265"

PN WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP01381.

XX 24-FEB-2000; 2000JP-0047405.

XX (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX WPI; 2001-582053/65.

DR P-PSDB; AAG65264, AAG65265.

XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -

XX Example 2; Page 58-123; 257pp; Japanese.

XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermitilis genome.

XX SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 534; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred. No. 5.2e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAACACACCTCATCACCGCGGAACCGCGGCTTCTCTACTCTCTCCGCGCGGCACC 60
Db 5143 GGAACACACCTCATCACCGCGGAACCGCGGCTTCTCTACTCTCTCCGCGCGGCACC 5202
QY 61 ACACCCACCAACCCCAACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 5203 ACACCCACCAACCCCAACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5262
QY 121 CACGACACACCTCAGCAGCAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
Db 5263 CACGACACACCTCAGCAGCAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5322
QY 181 TGCGACACCAAGCAACCCAGACCAACTCTCAACCACTCTCAACCACTCTCAACCACTCTCAAC 240
Db 5323 TGCGACACCAAGCAACCCAGACCAACTCTCAACCACTCTCTCAACCACTCTCTCAACCACTCT 5382
QY 241 CCCCTCACCACCGTCTATCCACACCGGAGGATCTCTCGAGCGCCACCTCTCAACCACTCTCT 300
Db 5383 CCCCTCACCACCGTCTATCCACACCGGAGGATCTCTCGAGCGCCACCTCTCAACCACTCTCT 5442
QY 301 ACCCCACCACTCAACAGCTCTCTCCGCGCAAGCCACAGCGCCACCTCTCTCTCTCTCTCT 360
Db 5443 ACCCCACCACTCAACAGCTCTCTCCGCGCAAGCCACAGCGCCACCTCTCTCTCTCTCTCT 5502
QY 361 CAACTCACCACCAACACACCCCTCTCAACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 5503 CAACTCACCACCAACACACCCCTCTCAACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5562
QY 421 TTCCGGCGCACCCGCGCAAGCCAACTAGCGCGCAGCAAGCCCTACTCTCTCTCTCTCTCTCTCT 480
Db 5563 TTCCGGCGCACCCGCGCAAGCCAACTAGCGCGCAGCAAGCCCTACTCTCTCTCTCTCTCTCTCT 5622
QY 481 CACGACCGGCACACCGCACCACTCTCCGCGCACAGCATCGCTTGGGCGACCTGG 534
Db 5623 CACGACCGGCACACCGCACCACTCTCCGCGCACAGCATCGCTTGGGCGACCTGG 5676

RESULT 4

AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.

XX AC AAZ58381;

XX 23-MAY-2000 (first entry)

DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.

XX Polyketide synthase; avermectin; insecticide; ss.

OS Streptomyces avermitilis.

PN WO200001827-A2.

XX 13-JAN-2000.

XX 06-JUL-1999; 99WO-GB02158.

XX 06-JUL-1998; 98GB-0014622.

XX (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.

PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
DR McArthur HAI;

XX WPI; 2000-182117/16.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI: 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermectilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 534; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 5.4e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAACCACTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 60
DB 5143 GGAACCACTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 5202
QY 61 ACCACCACTCAACCAACCAACCACTCTCTCTCACCAGCGGGAACCGGCGCTCGCCACCC 120
DB 5203 ACCACCACTCAACCAACCAACCACTCTCTCTCACCAGCGGGAACCGGCGCTCGCCACCC 5262
QY 121 CAGGCAACACCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 180
DB 5263 CAGGCAACACCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5322
QY 181 TCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 240
DB 5323 TCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5382
QY 241 CCCCTCACCACCGTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 300
DB 5383 CCCCTCACCACCGTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 5442
QY 301 ACCCCCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 360
DB 5443 ACCCCCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5502
QY 361 CAACTCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 420
DB 5503 CAACTCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5562
QY 421 TTGGGCGCACCGGCGCAAGCAACCACTACGCGGCAAGCAACCAACCAACCAACCAACCAACCA 480
DB 5563 TTGGGCGCACCGGCGCAAGCAACCACTACGCGGCAAGCAACCAACCAACCAACCAACCAACCA 5622
QY 481 CACCACCGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 534
DB 5623 CACCACCGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5676

RESULT 2

AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermectilis avermectin aglycone synthase DNA aveAI SEQ ID NO:1.

XX Streptomyces avermectilis; avermectin aglycone synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermectilis.
XX
FH Location/Qualifiers
FH 1..11919
FT /*tag= a
FT /note= "avermectin aglycone synthase protein"
FT CDS 1191..30690
FT /*tag= b
FT /note= "avermectin aglycone synthase protein".
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 95JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI: 2000-565458/52.
DR P-PSDB; AAB23749, AAB23750.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 2; Page 66-134; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 534; DB 21; Length 30690;
Best Local Similarity 100.0%; Pred. No. 5.2e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAACCACTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 60
DB 5143 GGAACCACTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 5202
QY 61 ACCACCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 120
DB 5203 ACCACCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5262
QY 121 CAGGCAACACCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 180
DB 5263 CAGGCAACACCTTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5322
QY 181 TCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 240
DB 5323 TCGGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5382
QY 241 CCCCTCACCACCGTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 300
DB 5383 CCCCTCACCACCGTCTATCACCAGCGGGAACCGGCGCTCGCCACCCACCTCACCACCACTTC 5442
QY 301 ACCCCCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 360
DB 5443 ACCCCCACTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5502
QY 361 CAACTCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 420
DB 5503 CAACTCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5562
QY 421 TTGGGCGCACCGGCGCAAGCAACCACTACGCGGCAAGCAACCAACCAACCAACCAACCAACCA 480
DB 5563 TTGGGCGCACCGGCGCAAGCAACCACTACGCGGCAAGCAACCAACCAACCAACCAACCAACCA 5622
QY 481 CACCACCGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 534
DB 5623 CACCACCGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5676

Result No.	Score	Query Match	Length	DB	ID	Description
1	534	100.0	11916	22	AAH79279	Streptomyces averm
2	534	100.0	30690	21	AAA92301	S. avermitilis ave
3	534	100.0	30690	22	AAH79277	Streptomyces averm
4	532.4	99.7	12381	21	AAZ58381	Streptomyces averm
5	266.8	50.0	125401	22	AAAD17186	Streptomyces nours
6	260.4	48.8	65140	22	AAAD17184	Streptomyces nours
7	257.4	48.2	4674	21	AAA44666	Nucleotide sequenc
8	257.4	48.2	4725	21	AAA44665	Nucleotide sequenc
9	257.4	48.2	4737	21	AAA44668	Nucleotide sequenc

THIS PAGE BLANK (uspto)

gene TDPAEVLVVTMDRGPHGAVGIIILEKKPSTWHLKLLATSCRIVSVFSGAGATILNLW
DQCARAGALHVLADFRDTRNRNMEIAYRFPAGFADSDPCVSEVAGASAAGVERLHLE
PSARPAPPTLTLLTAADIAPVTVSAAG"
complement (9895..10995)
/gene="fkbj"
CDS complement (9895..10995)
/gene="fkbj"
/notes="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="Fkbj"
/protein_id="AAF86388.1"
/db_xref="GI:9280390"
/translation="MPERDALLDLVDRAAEMDTSDELPRDLVRLGADGLICA
AEHGLGLGSGRENGEFTAHVSLCSLSIRSVMTSQGMAAVTORLGDAGORATFLKELT
SGKLAAGFSEORAGSDLSAMRTRVRLDGDTAVVDGKHVWTTAAAYADHLVVGLEQED
GSGAVVVPADTPGVVERVPKSGGCGRAAGHDLHLDQVVPAGAVLAGSGASLEPMLV
AASLAVGKSVANGCVGILKACRTAAVAHARTREQFRPLGDHQLVAGHIAIDLWTAEQ
IAARVCEYASDHWDEGSPENVPATILAKHVAERAAAGAAATRAQVILASAGAREGHVVE
RAYRDAKLMEIIEGSEMCRCVMLAQHALALPA"
gene complement (10988..11248)
/gene="fkbj"
CDS complement (10988..11248)
/gene="fkbj"
/note="ACP"
/codon_start=1
/transl_table=11
/product="Fkbj"
/protein_id="AAF86389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAHALVAFILKTRSDPVPVDRDLFABEGGLTSLFAM
ELVVYLEKTFDVTIAGPDLQLANFRVTVESNVALVHRLRAVDA"
gene complement (11245..12093)
/gene="fkbj"
CDS complement (11245..12093)
/gene="fkbj"
/notes="acyl CoA dehydrogenase"
/codon_start=1
/transl_table=11
/product="Fkbj"
/protein_id="AAF86390.1"

Query Match 45.7%; Score 244; DB 1; Length 77534;
Best Local Similarity 67.7%; Pred. No. 3.9e-22;
Matches 359; Conservative 0; Mismatches 165; Indels 6; Gaps 1;
QY 5 CGACCTCATCAGCGGGNACCGGGCCCTCGCCACCCACCTCACCCACCGCTCACCA 64
DB 61511 CCCTCCACACCAACACCCACCCACCCACCCACCCCTCAACCCCGAAGACGCGCATCA 61570
QY 65 CCACCAACCCACCAACCTCTCTCTACAGCGGACCGGCGCCCGCCACCGCCCGCAG 124
DB 61571 TCACCGGGGCTCCGGACCGCTCGCGGGATCTCTCGCGGCACCTGAACCCCGCCACA 61630
QY 125 CACAACACCTCACACCCCACTCAACAAAGGACATCCACCTCACCATCACCACTGGG 184
DB 61631 CCTACCTCTCTCCCGACCCCGCCCGGAGCCACCCCGGACCCCGCTCCCTGGG 61690
QY 185 ACACGAGACCCAGACCACTCAACAACTCTCAACACCATCCCCCGACACACCC 244
DB 61691 AGCTCGGCGACCCCGACCACTCGCCACCCCTCACCCACATCCCC-----CAACCCC 61744
QY 245 TCACACCGCTCATCCACAGCGAGGATCTCTGAGCGGCGCCACCTCACCAACCTCACCC 304
DB 61745 TCACCGGCATCTTCACACCGCGCCCGCCCTCGACGCGGATCTCTCCAGCCCTCACCC 61804
QY 305 CCACCCCACTCAACAACTCTCCGCGCCAAAGCCCAAGCGCCACCTCTCTCCACCAAC 364
DB 61805 CCGACCGCTCACACCGGCTCTCCACCCCAAGGCAACCGCGCTGGCACCTGGACACC 61864
QY 365 TCACCCCAACACACCCCGCTCAGCGCTTCTGCTCTACTCTCTCCGCGCGCGACCTTGG 424
DB 61865 TCACCCCAAAACCAACCCCTCACCCACTTCGCTCTACTCTACTCTCCAGCGCGCGCGCTCTCG 61924

QY 425 GCGCACCCGCGCAAGCAACTACGCGGACGCAACCGCTACCTCGACGCGCTCGCCACC 484
DB 61925 GCAGCCCCGGACAGGAATACTAGCGCGCGCGCAACCGCTTCTCGAGCGCTCGCCACC 61984
QY 485 ACCGCGACACCCACCAACCTCCCGCGCACACGATCGCTGGGGCACCTGG 534
DB 61985 ACCGCGACACCTCGGGCAACCGCGCACCTCCATCGCTGGGGCATGTGG 62034

Search completed: June 17, 2003, 17:55:34
Job time : 1038.31 secs

ascomyceticus (ATCC 14891) contains genes for biosynthesis of
unusual polyketide extender units
Gene 251 (1), 81-90 (2000)
MSLINE
PUBMED
20323220
10863099
REFERENCE
2 (bases 1 to 77534)
AUTHORS
Reeves, C.D., Wu, K., Chung, L., Revill, P. and Katz, L.
TITLE
Direct Submission
JOURNAL
Submitted (16-FEB-2000) Kosan Biosciences Inc, 3832 Bay Center
Place, Hayward, CA 94545, USA
FEATURES
Location/Qualifiers
1..77534
/organism="Streptomyces hygroscopicus var. ascomyceticus"
/db_xref="taxon:132248"
complement(412..1836)
/gene="fkbw"
complement(412..1836)
/gene="fkbw"
/codon_start=1
/transl_table=11
/product="fkbw"
/protein_id="AAF86380.1"
/db_xref="GI:9280382"
/translation="MVLPLGVGRPRRTGLALFLAAVAVLLPSPATARPAPDAPDG
HCLRSQRLVPAARQAQDCIDELTAGTAATGHTDPADWAGLTPKNLATPTGPGIG
IDGYPDSTTNNHNDHDAQFVIRLPDRWNGGLVAGTTCNREOYANDRAIDWVL
ARGYAYATKNTGTPAFYRDSRPGDAVAENMTLTOLTRAARASVAORVHRPLERT
LVYMSNGYLVRWQENHPYLDGGVDWEGTLWRAEGFNFTFLPALRDYPTFAAG
PAGQACREALLRAGFPDSEFLWPYHHEVYWDLTQVYREEFDGFGKEAGTFCY
APGTGDCADYATRSQVHAVERIALTGRIKGLITITQALSTCCCLPGGLRRVR
AVRQRPGRVRYRIEQTHTDALVDTPDRLRLPLLRCHRSAPTALESWLTGHRPP
ANHTVRPAGDAQSLLAECLDG"
complement(2020..3579)
/gene="fkbj"
complement(2020..3579)
/gene="fkbj"
/codon_start=1
/transl_table=11
/product="polyhydroxybutyrate depolymerase"
/protein_id="AAF86381.1"
/db_xref="GI:9280383"
/translation="WQPPPPRGILTLPLSSPPVGSLSRPRGRGVLTRLVAVVAVL
LGAALGTAPATAAAGLAKPLGLTADILTEVADFGTNPRLNMYVYRPSLPAEPVY
FALHCTDOAQYADNSGLLSFADRYGFLLVFAETTSNNANRCNFWFQSDNRRGQG
EASIRQWAHTVSAGADPORTYITGLSAGAMTSVMLATYPDVFOAGAVAGLPEG
CATDSSAYLCNPGTDLTADQWARRVRDGYPSNPGPRVAILWGHDKTTVAPRNAD
ELRDQWAVHVSQTPDTSVIGPNSRHEEYLAADGSAVAVENVRVPGIGTGPVDPG
TGAQCGSTGAAYFLDSCSYMITQFFLGSASDPGSLPAPSGLAATGATDTTISL
TWKPDGATDYAVHRDGAQITTSATTSYDTGLRAGTSHTYAARAADADGKAGPLSGA
VTAQTTGATVACWTAGNYAHVQGRATTSAGTYAKGSGQNGLYNTFTVTTLKESPT
GYPTANDTCp"
complement(3969..4496)
/gene="fkbR2"
complement(3969..4496)
/gene="fkbR2"
/note="transcription regulator"
/codon_start=1
/transl_table=11
/product="fkbR2"
/protein_id="AAF86382.1"
/db_xref="GI:9280384"
/translation="WVEGQGRYFEDFTGDDVYRHLGRVLTDSNMLTLTQNTAP
LHPDRHSAGTTWKGFLVDSITLALVTSQVTDLSOHVMANLGNDRVRLNPNPVEGD
TIYSQSEVLNTHRESARPDVGIISVNTIGYNQDGVIIITFERTLMVYRCHGPRFAAV
BPVWDRSRRAAGLH"
complement(4595..5488)
/gene="fkbR1"
complement(4595..5488)
/gene="fkbR1"
/note="transcription regulator"
/codon_start=1
/transl_table=11
/product="fkbR1"

protein_id="AAF86383.1"
/db_xref="GI:9280385"
/translation="MBIRELOWFTALAEHEHVTITAEIRLNIQSPTLSAIRRLERRVYG
VPLPDRHONELRLNKYGEFRAHRAISVSEGAQRITALIDPDTGTCTSPAGDGLHSYVG
TWLPDLAGYRRLAAGTTFELRGDADAADVDRNSRLDLGLTSPRPAGDDVEMTVP
RDEALCLLVPPGHRRLARRRVRTAELEAFVLEPVGFLGRQITDLRLCAAGFPVRV
LESTELSTRALVAAGLGVAVAPLTGSPGFGDHAVAPLDDPQARRTKVGVVTLAGGP
RAPVVARFYDYVRSAHAPHV"
5601..6818
/gene="fkbE"
5601..6818
/gene="fkbE"
/note="L-carnitine dehydratase"
/codon_start=1
/transl_table=11
/product="fkbE"
/protein_id="AAF86384.1"
/db_xref="GI:9280386"
/translation="MSEPHRPEQERPAGPLSGLLVVSLEQVAAPATRHLDLGR
VIKIERPGSGDLARGYDRTVRGMSHFVNLNRKESVQLDVRSPGEGNRHLHALVDRAD
VLQNLAAGAAGRLASATRSRGATEADHLRTYPATAVPAATADRKAYDLVQCEAGL
VSLTGTPTSPKVLGSLTADICAGMYAVSGILTALLKRARTGRGSQLSVSMLEALGEWM
GYABYTRYGCTAPARAGASHATIAPYGPETTRDGTINLGLONERERWASFCGVVLOR
PGLCDPRFGSNADRVARHYTELDALVSEVTGTLTGSELVARLESASITAYARQVTRSF
SEHPQLDRGRWAPFDPSPVGALEGLIPVTFHGEHPRRLGRVPELGEHTSVLAWLAA
PHSADREAGHAB"
6808..8052
/gene="fkbF"
6808..8052
/gene="fkbF"
/note="similar to Rhizobium trifolii MatC; dicarboxylate
carrier"
/codon_start=1
/transl_table=11
/product="fkbF"
/protein_id="AAF86385.1"
/db_xref="GI:9280387"
/translation="MPNELTGVLTILAAVFLLAGVRLNMGLLALVATPLLGVALDRT
PDEVLAGFPASMPFLVAVTFLGFIARVNGTVDMLVRVAVRAGVAGVAPVWFLGLA
ALLCATGAASPAVAIVAPISVAFVRRIDPLVAGLMAVNGAAGSAPGSGILGGIV
HSALBKHLVPSGGLLFACTFAFNLAVANVSLVGRRLPEHPLDDEDTDPTGDDPAS
RPGAEHMLTAMAAVLGTTVSLDSTGFLATLAAALLLFPRTSQATKEIAPHVY
LLVCGIVTYVALLQELGIVDSLGKMIAAITGTFLLAALVICVGGVWSAFATTTGLGA
LMPISGFLKSGAIGTTGMVMAALAAAATTVVDASPFSTNGALVANAPERLPRGVYQGL
LWMCAGVCALAPAAAWAAVVA"
8156..8824
/gene="fkbG"
8156..8824
/gene="fkbG"
/note="O-methyl transferase"
/codon_start=1
/transl_table=11
/product="fkbG"
/protein_id="AAF86386.1"
/db_xref="GI:9280388"
/translation="MANQITLSDTLAYVRKVSRLRDEVLRLRAQTAEPLGGVLPV
QAERGQFLPLVRLTGRQVLEIGTYGTYLCLARLAPGCVVTVCDMPKPEVGE
RYWEAGVADRIDVIRIGDARTVLGLDEAGAPESFDWIFIDKACYPAYBAALP
LVRRGGLIVDNTLFFGRVDEAVQDPDTVAVRELNALRDRDDRDVLAWLMTADGVTL
LRKp"
complement(8802..9884)
/gene="fkbH"
complement(8802..9884)
/gene="fkbH"
/codon_start=1
/transl_table=11
/product="fkbH"
/protein_id="AAF86387.1"
/db_xref="GI:9280389"
/translation="WTIVKLVMDLNTLWRGTLEDDEVVLTDEIRVITLDDRGI
LQAVANKNDHDLAWERLERLGVAFYVLARIGMGPKQSREIATLAFPTTIAFID
DQPARAEVAFHLPEVRCYPAEQANLLSLPEFPSPVSTVDSRRRLMYQGFARDQA
REAYSGPDEDFLRSLDLSMTAPAGEELSVEELTLRTSQNNATGVHYSOADLRACF


```
Qy 61 ACCACCCACCAACCCACCAACCTCTCTCTCACGACGCGAACCGGCCCCACACCCCC 120
Db 20052 GTACACACCCACGCGCGCGACCTCTCTCTCACGACGCGAACGACCCCAACGCCCC 20111
Qy 121 CAGGACAAACACTTACCAACCAACTTCAACAAAGGATCCACTTACCACTACCAACC 180
Db 20112 GCGCGCCACCAACTTACCAACCAACTTCAACAAAGGATCCACTTACCACTACCA 20171
Qy 181 TCGGACACCAACCCAGACCACTTCAACAAAGGATCCACTTACCACTACCAACAC 240
Db 20172 TCGGACCTTACGACGCGCGCGACCTTCAACAAAGGATCCACTTACCACTACCA 20222
Qy 241 CCCCTCACCAACCTTACCAACCGGACGATCTCTCGACGCGCGACCACTTACCAAC 300
Db 20223 CCCCTCACGCGCGTCTGTCACACCGCGGAGCGCTTCTCGGACACCACTTACCA 20282
Qy 301 ACCCGACCAACTCAACAGCTCTTCCGCGGCAAGCCAGCGCGCGACCACTTCTCC 360
Db 20283 AACCCGACACACTTACCAACCACTTCAACAAAGGATCCACTTACCACTACCA 20342
Qy 361 CAACCTCACCAACCAACCCCTTCTCGGCGCTTCTCTCTCTCTCTCTCTCTCTCT 420
Db 20343 GAACCTCACCAACCAACCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 20402
Qy 421 TTGCGCGACCGCGCAACCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Db 20403 CTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20462
Qy 481 CAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534
Db 20463 ACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20516

RESULT 14
AF357202 113193 bp DNA linear BCT 17-JUL-2001
LOCUS Streptomyces nodosus amphotericin biosynthetic gene cluster,
DEFINITION complete sequence.
ACCESSION AF357202
VERSION AF357202.1 GI:14794889
KEYWORDS Streptomyces nodosus.
SOURCE Streptomyces nodosus
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 113193)
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finnian, S.M. and O'Leary, M.
TITLE The amphotericin biosynthetic gene cluster from Streptomyces
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 113193)
AUTHORS Caffrey, P., Lynch, S.V., Flood, E.M., Finnian, S.M. and O'Leary, M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2001) Industrial Microbiology, University College
Dublin, Belfield, Dublin, Ireland
FEATURES
source 1. .113193
/organism="Streptomyces nodosus"
/db_xref="taxon:40318"
/complement(4. .1824)
/gene="amphg"
/complement(4. .1824)
/gene="amphg"
/notes="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="Amphg"
/protein_id="AAK73498.1"
/db_xref="GI:14794890"
/translation="MAPDVPEHEERSEQVRLAALLRPHRSVGLTAGVVG
LNAFGLGGRVTDIAGVGHGPGAFVDFGALGRLMLLLVIVVASVFLVQN
WLIVASVRLIHLRRAQEKARLPLRYFDRKPAKETLSRGTDVNDVQOQTQTLT
DLISVFSVLVIMLSLIIISPSLAGVMLLSVPVSGLLAAWISKRAQPOYAAQWSAGK
```

gene
CDS

gene
CDS

gene
CDS

```
LTATHEMAGHALVKAFDRRAEAEQRFDERNEAVYRAGSAQFASGAIPEVMFVAV
LVYAVAVGAKVNGSLTGDVQAFILYARQSPQITVIESVAGSLSEVGSQAVRV
FTLLDAPQEPKPDRLAVERGEVDFQVSPRYSPDTPLIEGLSLSEVGSQAVRV
GPSAGKTYVANLLMRFEYIDSGRIILDTGTAAAMNDDLRSRFGLVLQDTMLPKGTI
AENIAYSGPATRADIVAAATYADRIITLSQGYDTVLDDSGGVSAGSKQLITVA
RAFIAPAVILVLDATSSVTTRTELLIORMANTLURPGRTSFVIAHRUSTIRADIVV
MKSRIVEQGHDTQDIDAGAYARHAARADAPAAADVTG"
complement(1805. .3628)
/gene="amphH"
/complement(1805. .3628)
/notes="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="AmphH"
/protein_id="AAK73499.1"
/db_xref="GI:14794891"
/translation="MAPSVRMTPTFQRAAGRRPGCHSVLLRLMRLQRLPYAGSVVA
LVUHLVQIILGTLTLLPTLGAALIDEGVVRHDSRIGTIGATMAVVALVQIAALGAA
LGARTALGRDLRSVAFRRVLDFSAEIRGFSTLTSRVNDVQVQVQVQVQVQVQVQV
FVAAPLCLGSLVALLAQDVTALILVPMVLVAVCFGLLSRMAALYARLQOQLDRI
GRLLRERTGVVRVRSFARDAHEGERFTRTNEELGLSLGVRLIAVMLPSVLLMNL
FTLLGWGARRIDSGSMQIGALSAPLSYLSLLMSVVMAPFLNVPRAVCAERIT
EVLQETDVPPSPRPMAGQVVELVGAEFRTYPGAENAVLRDLSLTLRGERVAVL
GSTGSKTLLHLILRLVDVTAGVRIIGTVDRELDPSVMAAAYVQVPPRYLFAVT
ASNNRFGRPDADTELMEVLRIAGHEGFTRLGLDTEIAQGGTTVSGGQRQRLAIAR
ALLRRPAIYLPDDSFALDOSTEALAKALVPYTEGATVITVAQRVASVRDADRIVLL
DOGGIANTGTHALLRDSPTYREIALSORTREETAHGAGRS"
3840. .4874
/gene="amphDIII"
3840. .4874
/gene="amphDIII"
/notes="mycosamine biosynthesis protein"
/codon_start=1
/transl_table=11
/product="AmphDIII"
/db_xref="GI:14794892"
/translation="MPKRALITGITGODSGYLAHLLSQGVVGLIRGOANPKFRV
SRLASELSFVDGLMDQGSILVAVDKVQDPDEVNGLAISFPVMSQQAELVTEVNGMG
VLRVEAIRMVSGLSMERTAGTCQIRFYQASSEMFGKVAETPQRETTLPHRSPYG
AAKAGHPTTIRYRESFGMYAVSGLFNHESPRGQEFVTRKISLAVARIKLGQDLK
ALGNDAVRDYGADYVVRAMHLMQDQADPDYVIGTGEMHTVDRDAVAFAPFVGLQDL
KDYVVPDVLVRPAEVEVLCADSSQAQLGKPSVDFQELMRMMVDADLASVSRQNE
LDDLALLAHSW"
5042. .33574
/gene="amphI"
5042. .33574
/gene="amphI"
/notes="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/transl_table=11
/product="AmphI"
/db_xref="GI:14794893"
/translation="MDNEQKLRDYLKATADLRARRRVRGLESASQBPATVGMTCR
YPGGVSSPDLWRMVEAGENGVTPTDRGMDLEALASSPTSRGGLFADPEFADPF
GISPREAVMPDQORVULESWEAFERAGINPTSVKGSRTGVFGAMAQDVRVGPADG
AEGFOLTGNTGSLGISYTFGTGVPATVITACSSSLVAVHLATQALRAGCTLAL
AGGVTVMSGPGTFTIEMGRQSGSLVDGRCSFGDTADGTGWAEGVILVLERLSDAIRN
GREILAVVRGTAVNQDGSANGLTAPNGPSQQAQVIEQALVNARLSAGDIDVVEAHGTGT
TLGDPVEAQLLATYGOORDEDKPLLGLSVKSNISHTQAAAGVAGVIRKVMAMRHGVL
PRTLLADEPTRHVDMSQAGVRLTENTEMPAAGPRAAVSFGISGNTNHTIVEQAP
EPEDPADDDAPSTPAATVGVLPVLLSGRSEVRLAQAALLTGLCTGTPPADLAY
SLATRTTAFEHRAVLLASDLPELTGRLTAIEGDTDPVADLTVTGRTATRLAVLFT
CQGAQRGLAGRELAAAPFPFAAALDAALDFTPHLDVLRKVLKMGEDADRDLDRYAO
PALFAVEALYRLAESPEVKPDHLAGSHVGEIAAHAVGVSFLSDAAFLVAARGRLMO
ALPEGGAMVAQASEDEVAPLLAGEHDLVSLAANVGPVAVLSGDTTETVTELAARLTA
DGRKTSRLRVSHAPSHPLMPLDEFRNVFGLTLHSPVPSDVTGEPATVAOLTS
PDYVVDHVRQAVRFADGIDMLARHDVTAFLGPDPSVLISMAQNCNLDAGSDALTVA
LREGRPDHTITAAALAHQTGTALHWACFTGTGARRTDLPTTYTFFRRRWRPRAVOG
GAADLRSVGLGAHHPLLSAAVSLADSECALLTGRISLLSHPWLAHDHTVTRGATLLPGT
```

```

FQTAGPPVTAQLVQAGPETLPTTLTGVEVLGPIQLDALALDELITREAHFALFVLSPAD
INGPAHVAFOVANNRHAAGLPALASWGREELTSRERIALFDAATTVDDASVLAIRH
ELAASADVTCAGALRAAERGRALLDLVISEVGIITGGRPHVDRQSGFRDGTMTST
SAVELGRLLAAATGLALPATLAFDYPNPDLAAAYLGLRGFDTSAGAPTEPFTDPAAP
SGEPALVWACELPGVTSPEELWELSSGGDGIISGFTDRGMDLHLYHPDPDHPG
TSVSRGGFLHDGDFDAAFGISPREALAMPQOQRLLETTWEVLERAGIDFTSLRG
SRGTGAVVMYHDIYISRLKTVPELEGYSSGGSSVSGRVSYTFGFGPAVSVDTA
CSSSLVALHQAQRTGCTTAVAGVTVMASPSYIIFSRQGLSPGRCRCKSFAAD
ADCAWSEGAATLLVERLSDAQRGLHPVVRGTAVNOGDASNGLTAPNGSPQQRVI
ROALTNAQLTATDIDIIEAHGTGKLGDPLEAQAVLAMYQORDLPVWLGSLKSNIGH
TQAAGVAGVMKAVLSMRHGVMPKSLHIDEPTVDWAGTVELLAEARDWFGADERP
RAGVUSVSGTGNARHVLIEQAPPEPTESAEVPEPSAAGPSGPGVLELPGSGVP
WLSAQNETALDALKQLVETETGVQIAHSLITORSSTLPHRAVLVGSITREEFATA
LQOITRTPDGGKTVLVPFGQSQWGMQRLMETSPLYRDTIHEIDTALSJYIDWSV
TDLITTSQCAPTWDRVDVQPVLFMTTALATLWQHGLTPDAIIGHSGQIAAAYTA
GALTQDAARIITLRSKALLPLTGHCTMISVLAPPEKIOEIAADPOHWNTRIWIATAHNA
PNNTITIGTQALTHLTQQLSHRHRMRWQLPGVDFAGHSGHIDQLHNQKILAPITP
QKPTIPTSITNDITDLDAYMRNLQPVLPQPAINHLTNNGYITLITSPH
TLPFSQETNPNTTHTLHNQNDTHALLTALGHAHTHGHPIWHTLTPPTKTTPLP
TYPQHTRYLANDKPTVQDQATQLGLIPTPHPLVPATTTTAAETGATLTHGINPAH
PNTDHAVNGTPLLPGTALVDMALHAGDHTDHTLDELIIHTPIHTHTPTTQLTISP
PNTNTHLTIVHTRNSPDGPTAHATGTLTQHOHTDPDPQTPWPTDPAEPIPLDGTQY
RLATGLEGPTFGQLDALWRDSRLFADVRLPNDTHTSADGYGHIHPALLDAALHAL
VAVSGDEIRLFPFMTGVTLHATGATKLEWELWDSAGSASRAEDPSQPVVTVESL
ASRVTSPEGRASASGELXPETWAMSGAGTGGGAEFEYEVPTVGGDVLADTHR
RSTETLRIQOHLAADTTPLVVEATHDDLAGAAVMGLIRTQATEHPGRIIILIDTNDH
PITSKALPTLIASGEPOARIQDAILTLPRLTVVAPTEPAGVGEVTLITGGTGLSGSL
IARHLVTEHGVRLVLVSRQGPDPAGCATELAELOLGAQVRSQDITDRAALAAVLV
DEVPAGVHTAGALADATVNDLADALATTIAKAKAAMWHEHTQHGTLFWPV
SSLASVAGQANAYAAANSFDALATHRTTGLPGTSLAWSWERTSWSAHLTRAD
HRLRAGARLPTDEGLEFFPGDALVQVTVATFOLPAIGRGAHVPLLRGLVRA
PRAAVAGDGAEPWTERIATLPADARERSVLTSVIEQVAAQALGPPDPAEPQAF
KSLGFLTAVFERNRLSAMTGLQLPATLVFDPYPAAGAVTRYLLQVSPAPAAAGPDD
DDERELRLVERVPVAREFAGVLEALLKLAGPDPARTPDAEPEDEGLIDAMTEA
LIRHWMGTGA"
9637..14925
/notes="orf2"
/codon_start=1
/transl_table=1
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="WTNEQVLEALVRSAKENARLRENASLQEGAAEPVAVIGMACRL
PGCVSPESLMSAGSDVAFPTDRGMDLGLYDPPDHLGTAYVREGGFLDRAG
AFDPDPFGISPREALAMPQOQRLLETTWEVLERAGIDPTTLRSGRTGVFTGIMYQDY
VTRLDRFAELSGYLGSAASVAGSRVSYTFGFGPAVSLTACSSSLVALHLAVRP
ALRLLARVAGGVTVMSVPHVPSRQGLAFDARCKSPDASDGTVLSEVGGLLV
ERLSDALRUGHVPLAVVRGTAVNOGDASNGLTAPGSPQQRVIRDALTDARLTVADVD
LLEAHGTGRLGDPTEAQALIIATYQDRDQPVWLGSLKSNIGHTQAAAGVAGVIRKTM
AMRHGMPTRLHVSPTPNVDSAGEVRLLTWARDWPSDQRPRAAVSSFGISGTNA
HVVLQAPAAPDPVPTPPAGVPWLSAHTGPALSALEARLAHVEDEAPLGDVA
HTLNTTRATLPHRAVVCETREFEVVALRAVPTRPRGFLAVMTGGSQRPQMGRCQ
LYEHFPVQAALDEVFATLPGREVMFPDQOATLQRTDHAQIALFAPETALYRLWE
SWGLRDPMSGVHSGEITAAHVSGLTLPDAVHLVTTGTMQNIAPPGCMALYATDP
MPOILQPHNDHTIISIAINGPHATVLSGDRTHLHIAITQNTKTNLWNSHAFSPL
HGIITYLEIGPDTLTALARTLPTTHLIPTRRNNEVESTEALGRVFSVGHV
DRALLPTGRTSLTPTPQRDPWLHDAGGAEGVEGGLGTTDHPILLGAVTVVADT
GELLISLUSTSHWLTDHTVNGTIVPGTALIDIALHAAHEDHTTIVDELIVHTPL
ALHTPPSGTGAEDTQHRPIALFQSLRDTTGTWRTHTTGLTNSNTHAAELTSWSTP
DAHQDLTAYQADTGLHYGPAFQGLHTLYQORDNTLFAETLEPEAAGLPTGHLHP
ALLDAAHPAAVSAARDESQIOLPSWAGTVVHSTGATHRLVRDITDGTGRLRLAT
DTTGOPVLTISLTRYPLHPDOLHAAQPETNHLTYLWGSPTIPTTTTHGPERVTVPT
AGDDILADTHRTIATQTLRIQOHLAADTTPLVVEATHDDLAGARVWGLIRTAQTEHP
GRILLIDHTSHSLDASVAGSQPQARIDGAILTLPRLTSLVSDVTRASTREPRDPL
SMGVLVTGTTGGGLSLVRLVTEHGVLDLVLVSRQGPDPADGATELTNLQOQHARI
RTISCDLTATLTIIDIEGLTVGHTAGALADTTIDHLPDITTLTLPKPNPAW
WLHELTQOHDIALFVFSVAGVLAGNAGYAAANSFLDALIHTRRQGLTGPISLAW
QWOREGGMWTLTQADHDMTRAGIHGLTDAEGTTLFDLDRGLGAVAPVKHLQPT
LNRADTVAPLRLVRAVPRABOATVPAESWADRIALMPREERDAVIDLVRQASR
VLCHSDARIDPQAFKDTGFDLSLTAVERFNRTVATGURLPATLIDPHPTLNUADH
LIAQGLPGDGGSSPVEVVRDLERFERQLFDRPDDEVGTALAVELRSLILAHLDAPAT
LPAARTASAVQLTLEDASQSLELFSIDKEFRASN"

```

CDS

CDS

```

14964..21266
/notes="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="NADSKTLLDYLKWVTADLYRTRERLAEVESAEPIAIIAGMACR
PGGVTSPEALWELVSSGTDALSGFPDDRGDMEGFDFSDPAQGTSGVYMRGEGFLTNA
TGFDPAFGISPREAVAMPDQOQRLLETTWEVLERAGIDPTSLGSRCTGFTGLMANS
YTVTRVRIPELESVAGNGEASVASGRVSYTFGEGPAMSVDTACSSSLTALHLAVR
ALURSECPLAVAGVTVFATPRIFSDLRQRGLARDGRCKAFSTADGTLGSEGVGLL
LVERUSDARQLRGLSVLVRGTAVNOGDASNGLTAPNGPQQRVITKAALADARLTAAD
IDLIIEAHGTGKLGDPLEAQALIIATYQDRQEPQLWMSGLKSNIGHSQAAAGVAGIIKI
VWAMHGMPTKGLHIDEPTVDWAGTVELLAEARDWPDVDRPRRACVSGVSGT
NAHVLEOAPPAPEPTESAEVPEPSAAGPSGPGVLELPGSPVPMVLSAONETAKD
LARQATLVGDSVRDIAHSLI STRAALPHRAVVGTVSREEFATALQITGHTPDGKTV
LVPFGQSQWGMQRLMETSPLYRDTIHEIDTALSJYIDWSVTLITQSOPGAPTWDR
VDVIOQVLTFTMTALATLWQHGLTPDAIIGHSGQIEIAAYTAGALTIQDAARITTLR
SKALLPLTGHCTMISVLAPPTOIQIADPDHWNTRIWAHNAPNNTTITGDTQAITH
LTQQLSHRHRMRWQLPGVDFAGHSGHIDQLHNQKILAPITPQKPTIPFYGTTDNTW
ITDITLADAYMRNLQPVLPQPAINHLTNNGYITLITPTTPTTPTTPTTPTTPTTPTT
IHTLNSQNDTHALLTALGHAHTHGHPIWHTLTPPTKTTPLP
ADAVESAGLSRTDHPILGATTLTLAGHSETVLTGRLSLTHGHGWIADHANGVPTLPPT
ALVDMALHAGDHTDHTLDELIIHTPIHTHTPTTQLTITTPHTDNTLNLITTYRTAP
EQMTQHAAGTTLTRANAPTSPSETIWPPTNAEPIPLDGTQYORLATTGLEYPAGFQGLH
ALWRDSNRILLADVRLPDTIDNTNSTDCTDGYGHIHPALLDAALHALVAVSDGDIRL
PPAWGTVTLHATGATKLARELVIDDDGTLISLSDTSCQPVVTVSLTRPIDPAQLR
ARPTQDLYALDMLPWSGVAPAEFEQYVPRKGVDDVADLADTHRTITQTLTRVQOHL
AADTTPVVEATHDDLAGAAVMGLIRTQATEHPGRIIILIDTNDHPTSRDALASVVAS
GELPQARIRDGAILTLPRLTVVATVQALTPPADADAWHLDTSERGTLENITLTGSDTATR
NLLPCEVAVRAAGLNFRLVLIALGMYPGEMTGEAAGIIVLDIENDITLDTPGDVR
TGLPINALANTAVADRMTIPIPASWSNTDAAAFPIVLYTAYVGLHDLAALKPGSERVL
IHAGTGGVTAALQLAQLHGAEPVATAHTKWPHLRANGLDDHHLIATSRTPDFETFFH
TTTQCGHLVILNSLTDLTASLRLTPNGRFIEMGKTDVDRDTEVARHTPGVTYRA
FDIADAGDPDRIGAMLVELTGFERGALRGLPVRCDVRESVDAEFPAQARHIGKVY
LTIHTPDGTTILITGTCNLGLSLTAHLVTTTHGARDLILTSRQGNAPGACATLNLQJ
QHGARIITSCOLDTRTALTLLIDIEGLTVGHTAGALADTTIDHLPDITTLTPTTP
KANPAMWHELTQOHDIALFVFSVAGVLAGNAGYAAANSFLDALIHTRRQGLP
GTSLAWGWMQREGGTAHLTQADHDMTRAGIHGLTDAEGTTLFDLDRGLGAVAPV
KLUHTIARSGLVPLRALVRAARPTQAQAAQVQWADRVAAVABERRAVVLELV
RAQATLVHSGHAEHVSPTQAFKDSGFDLSLTAVERFNRTVATGURLPATLIDPHPTP
NTLADLQALAPAYTGPGLADLIGIEVTLRLTAEDEGRRPVAVARLKQWLSTLGAAG
PAADTLDLDAASAREIPDFLDGKAG"
21363..27176
/notes="orf4"
/codon_start=1
/transl_table=11
/product="PKS module 4"
/protein_id="AAC38064.1"
/db_xref="GI:2624951"
/translation="MANEDEFHLVYLKXAAADLRDARQOIQLEARDREPIAIVSNACR
LPGVSPESLWDLVLAARGDGVSAFFIDRGWMDLEGIYDPPDHPGTTTYVARGGFLDDA
AGFPDPAFGISPREALAMPQOQRLLETTWEVLERAGIDPAALRSGRTGVFTGTHND
YADASADDTQOLEGVLGNSSAASVAGSRVSYTFGFGPAVSLDTCSSSLVALHLAIO
ALRSECTTALTGGVTVLASVAPTEFSRQRGLADDGRCKAFADACSSSLVAGMAEGILL
LVERLSDAQRGLHVPILAVRGTAVNOGDASNGLTAPNGSPQQRVIRDALTNQALTATD
IDLVERHGTGKLGDPTEAQALIIATYQDRQOQPVRLGSLKSNIGHAQAAGVAGIITK
VWAMRHGMPTLHIDEPTVDWTTGTVLLETSDRWPTTDDHPHRAAVSFGVSGT
NLLHLEQATPTTPEPHASAEAPAGFPWMLISAHTEPALRELRLVTHREESVTDVAY
SALHGRALPCRAVMSITREDFAAELEALTCSGRASKVVFVFPQGSQWITGMGRGL
WESSAVPRDSVLRTRDALABFVDSAAAVALAGEPCTPDLDRVDVVQVPLFTVMVALN
VWRSYGVPEAVVCHSGOETIAAHVAGGLSIRDAARVPLDVRSLATALAGSGGMVAVQ
RSVAYVENLLQWEGRLSVAVVNGPEAVVVSQVAALEELIATEDRARRVADVAYASHS

```

Query Match 46.2%; Score 246.8; DB 1; Length 32870;

Best Local Similarity 68.0%; Pred. No. 2e-22;

Matches 363; Conservative 0; Mismatches 162; Indels 9; Gaps 1;

Qy 1 GGAACCACTTCATCACCGCGGACCGCGCCCTCGCCACCCACCTCACCACCCACCTC 60

DB 19992 GGAACCACTTCCTCATCACCGCGGACCGCGCAACCTCGCTCCCTCATCGCCACCCACCTC 20051

Query Match 47.3%; Score 252.6; DB 1; Length 28732;
Best Local Similarity 67.2%; Pred. No. 4e-23;
Matches 357; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 4 ACCACCTCATACCGCGGACCGCGGCGCTCGCCACCCACCTCTACCCACCCACCTCACC 63
DB 13094 ACCACCAACACACACACACACACCCACCCACCCACCCACCTCTACCCACCCACCGC 13153

QY 64 ACCACCAACCCACACACCTCTCTCTCACCAGCGAACCGCGCCCGCCACACCCCGCCAC 123
DB 13154 ATCTCTATCACCAGCGCGCTCGCGCACCTCTCGCGGCATCTCTCGCGCCACCTCAACCC 13213

QY 124 GCACACACCTCACCACCCACCTCTCAACAAACAAAGGATCTCCTCTACCATCACCCTGC 183
DB 13214 CCCTCACCCTACCTCTCTCTCGCGCACACACACCCACCCCGCCACCTCAACATC 13273

QY 184 GCACACAGCAACCCACAGCAACTCTCAACCAACTCTCTCAACACCATCTCCCGCCACCC 243
DB 13274 CCCTGCGACTCACCAGCCCGCCACCAATCACCAGGCGCTCACCACATCACCACCC 13333

QY 244 CTCACACCGCTCATCACCAGCGGATCTCTCGACGACGCGCCACCTCTACCAACCTCACC 303
DB 13334 CTCACCGGATCTTCCACACCGCGCGCACCCACGACGCGCCACCTCTACCAACCTCACC 13393

QY 304 CCCACCAACTCAACAGCTCTCTCGCGCCGCAAGCCCGCCACGCGCCACCTCTCTCCACCA 363
DB 13394 CCCCAACACTCACCACACCTCTCAACCCGCAAGCCCGCGCTCGGACCTCTCACCAC 13453

QY 364 CTCACCAACACACCCCTCTCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
DB 13454 CACACCCAAACACACCCCTCTCACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13513

QY 424 GCGGCAACCGCGGCAACGCAACTACGCGCGGACGCAACGCGCTCTCTCGACGCGCCCTC 483

Db 13514 GGCAGCCCGCGCCCAAGCCCACTAGCCCGCCGCCCAACGCTTCTCTCGACGCGCTCGCCACC 13573

QY 484 CACCGCCACACCCACCCACCTCCCGCCACCCACGAGCATCGCTCGGGGACCTGG 534

Db 13574 CACCGCCACACCCACCCCAAGGCAACCCCGCCACCCACCTCGCTCGGGGATGTGG 13624

RESULT 13

LOCUS AF007101 32870 bp DNA linear BCT 13-JAN-1998

DEFINITION Streptomyces hygroscopicus putative pteridine-dependent dioxigenase, PKS modules 1,2,3 and 4, and putative regulatory protein genes, complete cds and putative hydroxylase genes, partial cds.

ACCESSION AF007101

VERSION AF007101.1 GI:2624946

KEYWORDS Streptomyces hygroscopicus.

SOURCE Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

ORGANISM 1 (bases 1 to 32870)

REFERENCE Ruan, X., Stassi, D., Lax, S. A. and Katz, L.

AUTHORS A second type-1 PKS gene cluster isolated from Streptomyces

TITLE hygroscopicus ATCC 29253, a rapamycin-producing strain

JOURNAL Gene 203 (1), 1-9 (1997).

MEDLINE 98085969

PUBMED 9426000

REFERENCE 2 (bases 1 to 32870)

AUTHORS Ruan, X. and Lax, S.

TITLE Direct Submission

JOURNAL Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott

FEATURES

source Location/Qualifiers

1..32870

/organism="Streptomyces hygroscopicus"

/strain="ATCC 29253"

/db_xref="taxon:1912"

complement(108..1130)

/note="orf5; similar to RapK"

/codon_start=1

/transl_table=11

/product="putative pteridine-dependent dioxigenase"

/protein_id="AAC38060.1"

/db_xref="GI:2624947"

/translation="MNPSSLVNLGLTSYFENGRAVVPVPPGRNVLGVNYSVCEYPT

LDHGYPELEINNVAPTAEPFAEVVMTDAESHEGRDGTITYAHDGYFFCFACGVPPTGR

YTEATRAAYVTMFLLEEFYGSVFRMNFIDINRDNAGMEVYRDFCRGAEAFEPQ

CRLEPDPAPATGIGSGGGIAPYLLACRSGGHVHENPROVPAYHYPKRYGPRAPF

ARATYLPRAADGVGGGVFSGTASVLGHETAHEGDLVKQCLALENLELVISGGLA

AHGISAGHGLTALRNKIVYVRSEDVPAVREICREAFSPDADIVLTVDVCRSDLLVE

IEGVVM"

1454..9625

/note="orf1; loading domain"

/codon_start=1

/transl_table=11

/product="PKS module 1"

/protein_id="AAC38061.1"

/db_xref="GI:2624948"

/translation="MLRRELRIPLPELKAHAARRGSQTAFADRRRGVSVAQLEARTR

DSAGHIAQLRLPGDRAIYLCNVCETIYESYLAIRASAIQVVPINPHSSTABLYLLD

RLAARVIVTDHTRLPQVRAERPHLTVVVTGDBESGAVPFDALAEETPOQPARDDL

GLDIAWMLYSGTTGKPGVLSGTORSCLSWVAASSAGVLGSETDRVLPVPLPHSL

GHITCVLSVTANGATARITEAFDEBELLTLRABEFTFLAGVTMYHYLLDARDDGL

SLPNRLVCSAGAIATATRAEPQETFGVPLLDCTGCTETGLTANNPNTQAEQSC

GLPVGPLTVRIVPDETHESVETGEEIHWGPGSLMVGHYNQNPATEALPRGMYRTG

DLGRDDLLGYHTITGRSELIRSGENIHPTVEQVLLRVFGVDAAVVSAHPTLGE

VPVAFVPAADGFPDVEFAACRHLAVFKVPEELYEIGSIPTGSGGIKRLAQAQSP

ARLNAVSSGESPELRLFWTAVPVAGPDGQDATSAHRAGGSVTAGLEAGCAVDDC

PRGSJDEVLRAVRLTDRSGRAVLTGHADHPAQAAVWGLVSMRADADPHADPVLV

DADGGEATAALASVALSQPRALVRGGEVLPRLARAGSGQRVSRPDPGTVLVFGAD

QEQAELAGHLVAAYGIRHVRLLRTAVPDLVRLTELGAASVTQTVPRAQGRASLAEA

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

Eur. J. Biochem. 256 (3), 528-534 (1998)
98451508
9780228
2 (bases 1 to 28732)
Motamedi, H.
Direct Submission
Submitted (04-AUG-1998) Molecular Pharmacology/Immunology, Merck
Research Laboratories, P.O.Box 2000 Bldg. RY80M-213, Rahway, NJ
07065, USA

FEATURES

Location/Qualifiers

1..28732
/organism="Streptomyces sp. MA6548"
/strain="MA6548"
/db_xref="taxon:82632"
complement (61..4575)
/gene="fkbp"
complement (61..4575)
/gene="fkbp"
/codon_start=1
/transl_table=11
/product="FK506 peptide synthetase"
/protein_id="AAC68816.1"
/db_xref="GI:3798625"
/translation="MTSRGRPGTTLRGSSPKKRNSTNREIIVAADAPPLASFAQR
INFLQDDPSNAYNLPLVRLTRGLDPAALLERALTIVARHEALRTVPDAPGEPVQ
RVLPAPRILRHARADGEGQAARLVHDEIAAPFDLSGLIRALLIRVGGDDHVLAVT
VHHAAWGSGFLGLLEAAHCTALRDSRPVLPPLRYADFAAERRELITGAGLE
KELAYWRDLRGAPVLDLPTDRPRPVADADAGLVEMRPPAEALAAVRRIGRDAGTS
AFWTLIAFQLVLAHAGTRDVLVGTVPVAVNRTRTEFEGVIGNFVTLAQRDLSGDP
FRELDRCATAADAPAHADLPFENLIEVLPDRDLSVNFVQVLLQVRDASAPAL
PGVTAAPGTGKWFTRFDLEFHYBEADGALTGELLYSRALFDESRIITRLDEFTAVL
RAATAAPGTLSQLPQAEATPTPEISNDTARELPVDTLPELLARYAARTPDAAVTD
PCTSYTGELDRANRAHLRLARGTIVTGIRGICADRGVDLVGIVGLKAGAAVYV
LOPEHPERTDFLADARLTAVVAHETVHRFSGVPHVYVDPDPLGRQPDPAHLTL
GRDSLAYLYTSGTGKPKAVLMPGVSANVLLWQERTWGRPEASRTVQVTFATFDYS
VQEIFSALLGTYLTPDDEARDFPGLARWDEQALITRIAYTAVLRALVLEDPHSD
QUSALRQCGGEALADRLRELCHRPPLRHVNHGPAESQLITGLTPADPDTPW
AAATPGRTIDNTRIHLLDARLPDGMFQGLCVAGIGLARGYLARPELTAEWVSGG
ATEERMYTLGLARAPGGLDFLGRIDQVKIRIGIRIPEGEVENLAEADARVAHA
VSVRDPKEGFLAAVVPVTPAGHGLAASRLGLADRLPAYLVPSAVVMVMDALPRT
SLLATRVSRIRAEIGADVALTKLDFARTPAALARAADGAPPAALPILSSPCAGP
ITAAQEQMLHSGSLAAPSITYVAPYFRLGRLDRLDALTRITARHEPLRTGFR
GREQVQPPAAVPEVSGVGAARARAEIARLPFDLVAGPLLRALLPLGAE
HYLLMLHAGDGSFOLLVRELGSAPDLVSPVDPVWARVERVPAVVAARENDAYW
RQFGGACAPLPTVPRGTAAPDCGRATLTLDDATVAAREVADARNATWHETVLGA
FALVVAEADTDVILVATPPADRGVAGTDHLIGLFAKVLALRIDVSGSTGGTSPFAEV
LRVHATMTGAHQSVFYSAMRAADPALPPTPVPSFQLISALSTELRLPGLRTEFPV
VTVTDEINGELSVNLPDGGHTVSGAVVHDAALLDRATVDDLLTRVESTLRAAGDLT
VRVTCQVGSSE"

complement (4572..5570)
/gene="fkbo"
complement (4572..5570)
/gene="fkbo"
/codon_start=1
/transl_table=11
/product="FK506 oxidase"
/protein_id="AAC68817.1"
/db_xref="GI:3798626"
/translation="WSAPVAAPYCFEIKIGSPDLBEGDETVLGVIEHGTGHTGASLVGD
VPHLATHTRDEEAPFAEVMHAQPPVPSGTDGDIAMARTDEYLFQGVGRVPSGRYADA
VSALYTRISGLTRSLGHPHLLARTWNYVSGINSANADGLEVDYDFCVGRAQALDEGID
PASMAAATGITHGGGIIACAFIAARGGVYVNIENPAVLTAHYPAAYGPRSPVPRAT
WLPDPRGLFVSAATGILGHETTHGGDVTCQCEVALDNIAVVAENLRHGIIRGH
LLADYVDLKVYVRREDLIAANRVCAARLSSTATVAVLHTDIARGDLLLVEIGVVR"
5798..28528
/gene="fkbb"
5798..28528
/gene="fkbb"
/function="FK506 polyketide backbone synthesis"
/note="initiation and first four elongation steps"
/codon_start=1
/transl_table=11

gene

CDS

gene

CDS

/product="FK506 polyketide synthase"

/protein_id="AAC68815.1"

/db_xref="GI:3798624"

/translation="MARAESIRPLHLLRVAERHGRVAVYDSSRAVTVYGRLOLRGT
BIAGHLAASGVGRDVRVLLIGNRIETIEVLAARAANAVALPLNPDAGDAELAHPLT
DSGATVLDVDDHLQVRRVGTATVLAGRRPCVAYEDLADLPPRPFRPPRDDIIGLD
PEAWMLVTSGTTRPKGVSAQSLGWSATCDVPSWELTDEDELLMFAPLPHSGHH
LCLLAVLTGASARILGFGVARDLDALEHPCTVILVGYPTMYTLLGASVGEPOPTRA
LRVALVAGTSPASLTAGFAAAGVPLDFTYCTETTGSLTANTUTDRVPCSGCLPV
PCLSLRFVDPTGADVRREGELWASGSLMGLYHAQTEATAOVLADGWETGDLAR
QAEYTHVITGRVKELIIRGGENHPREIESVAQVPGVRDAAAAGRPHPLVGEIPLV
YVPEGPVPADAILAECRRQLAYFPDEILHVTTPRTASGRVRRGELAGLPAARLV
GTGGEAALCELWRELPDTSQVQAVTVTRAVVIGFAEVPDADQALRWDEALVRQ
AAEPGSFVLDTDDTDANNDEADELVAALAAAGVPRVLAAGREGVAVPRLVRADT
TPLPASOWALRPASGTLRLDPAVPSDAPRPPAAAGVEVRVRAAGLFRDVLIALGT
YPDGAMGGEAAGIVTEVPGVDLVPCHAVGLVQDAPRBSVVAADRILVARI PRGWS
FPVAAVPIVATANYGLVADAGVRPGQKVLIIHAATGGVSAARQIAHMGADVATTA
SPAKHLLADGFDHADVANSKTTAFADFPVVDVNLSTGEFIDASVELLAPGGRF
IEMGKTDIRHDVQPPFDLSVDYDPAARLEILLLELDFDRDELRLPVRPMDIRRADA
FGMMSQARTKGTVLTPRPLGPDAPVLVTGEGADAVAGLDRDETGRTHVSSAAEAE
RABEGEGVNASTLAHVLGPDAGADAVARGASEGLPIVTVHGLPGIEQALTCGLI
EAALEGSGSVYVTRVGSAGARAEAMAGATPPLISLACTGAEAPGAGPAGWSRLAA
ADARENVLILVDRSVATVGLPGAERCADPRTFRENGLSLTTVEFANTVAARTGL
RVPASTAFDHPTRPSFAAHLGASVAAAPAAKALSGSEPVAIVGMARLPPGGVTSPEGL
WLVASGTDATTEFPVDRGDVDELVDPPDAVGRSTTRHGGFTTEATGFAAFFGIS
PNEALAMPQQLVLETSWAEFAHAGIVPDTLRESDTGVFMGAHQFGAGRLGGLG
VTATQTSVLSGLSVFYVGLQGPVATVDTACSSSLVALHQAQALSGSCSLALAGVT
VMATPGSVFESRORGISPDGRCAFAADAAGTGAEGCVGLVVERQSDAERNCHTVL
AVRGSANQOGASNGLSAPNGVQAQVRVIRQALANAGLSGADVDVAEHGTTGTVLGPD
I EAQVAVLSQOEREPEPLLLGSVKNVGTQAAAGVGIKMMWAMRHGLLPKTLHVD
EPSSHWSAGAVEVETRETPESGGARRAGSVSGVSTNAHVLGVEAESSVRDGE
GSSGLPLPVSAARTESLALOVERLGEYVRGAGDAVADGALVRGVRADRAVLGGE
S AVAGA VEGARTVFPVPGQSWGMGRELKMGASEVFAARMRECAALEPHTGWDL
DVLGAUVVDEVEVLQASWAVASLALHQAAGVVPDPAVGHSGQEAIAACVAGALS
LEDARVVALRSQALAAKLAGKAWASTAVPASAVETVEGVVIAHNGPSTVAGDP
AAVERVLAEAGVRVRIADVYASHTPHVEALEADALEGITSPTSPVPMWSTV
DSGWTFEPDGYWYRNLPQVMDATVSELDSGLFIECSAHPVLLPALDDOVTASL
RTDDGWRDFRLLAALAAQATQGVDMVTLLIEPAPHRVLDLPTYPDKRXYLQAPVAT
GAGTIGFLSSAAVLPQSDGVLTKLSLATHPMLADHTVOEVLHCTAFLELVVRA
GDEVCOTIDELVLETPHSLPASGAVDLTVTDQDDTGHPRVSVHARPACTDITWRH
ATGILGRTDTPDITNSFPQWPTGAQSLDEFTERLAGAGTYTGAQGLRAARAA
GDTVVAALDGAADVDRFVHPALLDAALHGRDADGDELELFPFSWTVGRHLHAGAA
AVRABERGPAVAVADLDREPVSDALVLRPSVAPSNGLLGLEWLEFVABAHY
DADLEPEGYTLITATHPDDDDPTNPNTPTRTTQTTTTLTALQHLHLEWLPVATLI
VHTTTPPGAATGLTRTAONEHPGRILHIEHTHPPLTQLTTLHQLHRLTNNT
LTHPLLPITHTNTTTTTNTPLNPHAILITGGSGTLAGILARLHPLTYLLSR
TPPTPTTHIPCDLTDPTQITQALTHIPQPLTHFTHAATHADATUTNTPQHLTT
TLQPKADAAHHLHHTNQPLTHFVLYSSAAATLGSPQANYAANAFDLALATHTT
CQGPATTIAGMWHHTTTLSQLTDTTERQVDRGFRPLTEAGNHFLDASLATDPV
VAALPABPARRRERRTARTEGSDROLLAVVCVATAVLGHADASEIGAATPKDLG
IDSLSGILRLNSLAEATGRLSATVFDHPTPDALAAELAEGLDDSDSQGAAAAAA
PGTVPTVAAGDEPLAIVAMACRMPGVDTPEDLWSLVESGDDAITTEPTDRGMDLAA
LYDPPDAIGKVSVRHGFSGAADFAPFGISPREALAMDPOORLVLEVSWAEAFER
AGILPASVRGSDAGVFMGATQGGYAGVDLGGFGATGTTTSLVLSGRISYYLGEPSV
ADTACSSSLVALHQAARSLRSGCSIALVGGVTVMATTTVTFBFSRQRLAPDGRK
AFADTADTSPAEAGVLI VERLSDATRLGHVTVVVRGVSNDGSAENGSLAPNPA
QOVRERALDDAGLPGDIDAHEAGTGLGDPLEAQLAALAAVGLDOLHPLLLGSLK
SNRFGHTQAAGVAGVIVKMLAMRHGVLPRLTHVDEPSRHDVGDVRLLRNEPVRT
GVRGVSFGISGANVAVHLEAGPAAAPADPAADPVPDVPVVPVVPVVPVVPVVPV
AGRLTATAAAGIHSIATTRAMHRAVVPARDAEFARFGEVPGVGTADVTDT
RAVFPVPCQSQMDMGAEALITATEPVFARRLGECAALAPYTGWLLDVLVARRCPAE
LDRVYVQPASFAMWALAEVRAHGAAPAAVGHSGQEAACAVAGLUDADAABELL
ALRSVLATRAHGGVSVPPADPAAANAGRLVAAQVINGVAPASIVAGLUDADAABELL
AATPHARRIADVASHTAHVESIRGALLDALADLTGCAPEIPFFSTVDEMLDPADA
AYVDNVKRCVPRFAGAAARLAEHGRVFEASPHPLVTLTADLADHPNTAVTGTLR
RGDGAERFTRSLAEHLWVRGVPSVFPFELRGVLPPTYPRFRDWRVWDADPAGTSGHP
LLGSLVERADCEGAFATLFSRQRPQPLADHEVDGRIIVPGSALVLELAEAGARLGT
TIAELTIVAVVYDEERDTBIQLTVA TEVSRRRAVRLHSRTGAGTPTVGTALGASDT
RDPPEVHPADAEVDLTGFDYDPLPI SYGPAFRAMTAMWAGQGRASVYVPLGQLTD
ARYGLHPLVLDAAVHALGTLFADERRRLAFSWSGVRILHAKAATITQRVILLRADGDTI
RIALDHEGAPVLDVGLTVRAAPFTEALFEVAVWSVPASVPGWTHVLLADPBGDVI
EGDAGEGDVGESPPPVVLAPEFGDPDPSGPRARELGLLAVTRVTLAAPRAES
RIVVVVTRTGDPQAEGALGGLVRTAGTGHPRVGRVTEADEITPATVEAGLAAGDESHVR

WAVAVSLAALWQAHGVSDDAVIGHSGEIAAACVAGALSLEDAARIVARLSQTTAAHL
 AGRGAMASITATPATVETVEGVVVAARNGPESTVAVGDPSSAVERVLARYEAGRVRR
 TAVDYVSTPHVEALQEOALADVLGDI TSASPVMWSTVDCGWTEVPAGDDYVWNL
 QPVANDTALGELDSIFTECSAHPVLLPALDQERTVASLRDGDGNDRLTALAQAWT
 QGADVDTTLLIAPADRLDULPTYPFDHKKRYWIEATGAADLTALGLDITAHPLAATA
 ALPTNGLTVLIRISLRTHPLADHTI QDVLPLPGTAFLELVMRAGDVGCDTIDELY
 1BTPLTLPVTGAVDITVTVEQADATGHRPVSVHARPAGDTWTRHATGTLGRTEDTAP
 ISFTSPQWPTTALDVTVEFYAQAALGYREVPFTFGLRAARAGDITVYAEVALPED
 RAADADHGVHALLDAALQSGLLMLSDGEGVQLPFSWGHVGFHATGATSLRAVI
 VPGDGLRLRAADSENHPVATIDALVTRPPADLSPDPLRVGVWTAVPVPAEVRSDI
 DVLMLSDADDPIQETRDUTTRVLDALLTTEGLTIVQVTGLAANAAGLVRTAQSEQ
 PQRFLETDPGEVDGACDMIAALGEPRLRLGRNFEARLVRANAALTPDPDGSW
 QLRQSGTSLDDVAIVPTGEPDRPLGAGEVRIAVRAAGLFRDDTVALGVVADDRPLG
 SEAGVTLTGEGVHDLGAPVGLMAGFGVPAITDRRLGRMPDGTWTFPOQASV
 TVATTTWGLDLAELRPGCEKVLHAAATGVGAAGVQIARHLDDADVYATASTPKOHL
 DLEPHIADSRITATADFPADPVNLSTGLPDDASAOQLLTPGCRFTEMKTDIRHH
 VQOPFDLVAGDPDQOQIITQGLDLFANGTLHPVPHAWDVQAREAFGWSSGRHTG
 KLVLPVRLPDPEGAVITGGSGTLFAGLHRLDHPHTYLLSRTPPTTPTGTHLPD
 VSDPTQLATALAQIIPQPLTGI PHTAATLDDALLHNLTPQRI DTVLPQKADAAMHLHH
 THHATDLAAFLVYSAVAGVMGSPGNGYVANAFLDALAEHROAMGLPAOSLANGMW
 TDAITTPGDRGMDLDALYDPDPAVGRAYSIRGGFLGEGAEFADAFDIPSRESLGM
 DPQORLLTWEATEIRGRINPASLHGREIVGVYGAAGYGLGAEDETNATIGST
 SLLSGLLAIVLLEGPAVTDTACSSLLVALHLACQRLRGLGECALAGSVLSPA
 AFVBSRQGLAADRCRKSFGSGADGTTWAEVGVVILVLEQLSDAERLGHVTLVVRGS
 AVTSGASNTLAPNGLAQORVIRKALAAAGTAAADVLVEGHGTGLRGDVEADAL
 LATYQNGREPWMLGSLKSNIGHAKAAAGVAVKITVQAIGATMPRTLHADDEPSAV
 DWTAGQSVLLTGNRPWMTSVPERPSPRSGAGRTL TSSLSNTSVRSRRPDLARNP
 GRCLGCPRLPPLPCGRPPVCAI IWRPSGDPPLDIGNAQT SRAQFTHRAAEAAAP
 DGFRAGLSDVAGAEAPSSAEPPRSARFLFDGOGAORVGNRGLHGFVPVFAWMD
 EYSDAFGRHLESPTDVFHENGDLNHDITLYAQLVLTLEVALLLHLEHWGRVPDVV
 GHSVEVTAAYAGVLTADATTLVARGRALRPPGAMTAVEGSPAEVGTATDIDI
 AAVGPSAVLVTPDDVAADFEREWAARAKRLDVGHSFHSRVDGALDGFVGLIE
 SLAFGAARLVSTTTGDRDAAGDLATPEHRLHARRPVLVADARELADLVGNMFVAV
 GPSGLASASNTSGSAGTYHVLRRPARGESAAALTA VAEHLHAGAPVDLVAALVAG
 RPVDLPVYFQHRYSWLA PAVGSGSPATVPDTPTEPDPLTVAEIVRRRAAALIG
 IADPGDVADATTFPAGLFPDS LAVQRLRNLQTLTAATGLDPLTAVLFDHDPFSALTAYLEE
 QGDGADDDAPFTVLALLMDSLDAADTAATPAPERAAATADJLDLKLSTWTKDHR"
 46. .5740
 /gene="fkba"
 /function="approx. span of module 7"
 136. .1345
 /gene="fkba"
 /function="approx. span of beta-ketoacyl ACP synthase
 domain of module 7"
 1609. .2413
 /gene="fkba"
 /function="approx. span of acyltransferase domain module
 7"
 2509. .3028
 /gene="fkba"
 /function="approx. span of dehydratase domain of module 7"
 3668. .4699
 /gene="fkba"
 /function="approx. span of enoyl reductase domain of
 module 7"
 5521. .5716
 /gene="fkba"
 /function="approx. span of acyl carrier protein domain of
 module 7"
 5741. .10497
 /gene="fkba"
 /function="approx. span of module 8"
 5770. .7345
 /gene="fkba"
 /function="approx. span of beta-ketoacyl ACP synthase
 domain of module 8"
 7354. .8176
 /gene="fkba"
 /function="approx. span of acyltransferase domain of

module 8"
 /gene="fkba"
 /function="approx. span of non-functional dehydratase
 domain of module 8"
 9439. .10045
 /gene="fkba"
 /function="approx. span of ketoreductase domain of module
 8"
 10252. .10471
 /gene="fkba"
 /function="approx. span of acylcarrier protein domain of
 module 8"
 10498. .16234
 Query Match 48.2%; Score 257.4; DB 1; Length 19791;
 Best Local Similarity 67.8%; Pred. No. 1.1e-23;
 Matches 360; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
 QY 4 ACCACCTCATCACCAGCGGGAACCGGGCCCTCGCCACCCACCTCACCCACCACTCAC 63
 DB 9400 ACCACCAACACACCAACACACACACACACACACACACACACACACACACACAC 9459
 QY 64 ACCACCAACACACACACACACACCTCTCTCTCAACAGCGGAAACCGGCCCAACAC 123
 DB 9460 ATCTCTCATCAGCGGGCGCTCGGCAACCTCTCGCGGCATCTCTCGCGGCACCTCA 9519
 QY 124 GCACCAACCTTACACACCACTCCACAAACTCCACAAAGGACATCCACCTCACCACTGC 183
 DB 9520 CCCCACACTTACTCTCTCTCGGCACACACACACACACACACACACACACACATC 9579
 QY 184 GACACCAACACACAGACAACTCTCAACAACTCTCTCAACACATCCCCCCCCCAACACCC 243
 DB 9580 CCCTCGACCTTACCGACCCCAACCAATACCCCAAGCCCTCACCCACATACCAACCC 9639
 QY 244 CTCACACCGTCATCCACACCGCAGGATCTCTCGAGAGCGCACCTTCCACCAACCTCAC 303
 DB 9640 CTCACCGGATCTTCCACACCGCGGCCACCTCTCGACGACGCGCACCTTCCACCAACCTCAC 9699
 QY 304 CCCACCACTTCAACAACTCTCTCGCGCAAGGACACGCGCCACACGCGCCACCTCTCCACCA 363
 DB 9700 CCCCACACTTACACACACCTCTCTCAACACACCTCTCTCAACACACACACACACAC 9759
 QY 364 CTCACCAACACACACACCTCTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
 DB 9760 CACACCGAACCAACACCTTACCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9819
 QY 424 GCGCACCGCGCAAGCAACTAGCGCGAGCCAAACGCTTACCTCGAGCGCTTCCGCGCCAC 483
 DB 9820 GCGACCGCGCGCAAGCAACTAGCGCGCGCAACGCTTCTCTCGAGCGCTTCCGCGCCAC 9879
 QY 484 CACCGCCACACCCACCT 534
 DB 9880 CACCGCCACACCCACAGCAACCCCGCCACCACTCTCTCTCTCTCTCTCTCTCTCTCT 9930
 RESULT 12
 AF082100
 LOCUS
 DEFINITION
 Streptomyces sp. MA6548 FK506 peptide synthetase (fkbp), FK506
 oxidase (fkbo), and FK506 polyketide synthase (fkbb) genes,
 complete cds.
 ACCESSION AF082100
 VERSION AF082100.1
 KEYWORDS GI:3798623
 SOURCE
 ORGANISM
 Streptomyces sp. MA6548.
 Streptomyces sp. MA6548
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 Motamedi, H. and Shafiee, A.
 1 (bases 1 to 28732)
 The biosynthetic gene cluster for the macrolactone ring of the
 immunosuppressant FK506

Db 39661 GTACCGAGCAGCGCGCGCCACCTGCTGCTGGCGCGCGCGCGCGCGCGCGCC 39720
Qy 121 CACGACACACCTCACCACCCCACTCCAAACAAAGGATCCACCTCACCATCACCACC 180
Db 39721 GCGCGCGCGCACTCCACCGCGCACTGACCGCGCTGGCGCGCGAGGTACCGTGGCGGC 39780
Qy 181 TCGACACACGAGCAACCCAGACCACTCCAAACAACTCTCAACACCACTCCCGCCCAACAC 240
Db 39781 TGGAGCTGCGCGACCGACCGCGCTCGCGCGCTGCTCGCCACCGTGGCGCGAACAC 39840
Qy 241 CCGCTCACCACCGTATCAGACCGGAGGATCTCGAGAGCGCAACCTCACCACCTC 300
Db 39841 CCGCTCACCAGCTGCTGACACCGCGCGCTCTGAGACGCGCAACCTCACCAGCTG 39900
Qy 301 ACCCGCACCACTCAACAGCTCTCGCGCGCAAGCCACAGCGCGCCACCTCTCCAC 360
Db 39901 AACCGGACCGCTCGCCACCGTCTACGCGCCCAAGTGGAGCGCGCTGGCACTTGCAC 39960
Qy 361 CAACTCACCACCAACACCCCGCTCACCAGCTTGTCTCTACTCTCTCGCGCGCGCCACC 420
Db 39961 GACCTCACCAGCGACCTCGACCTGGCGCGTTCGTCTCTACTCTCCACCGCGCGCTC 40020
Qy 421 TTGGCGGACCGCGCAAGCACTAGCGCGGAGCAAGCGCTACCTCAGCGCTCGCC 480
Db 40021 ATGGCGGACCGCGCGCGCACTACGCGCGCGCAACACCTTCTCGACGCGCTCGCC 40080
Qy 481 CACCACCGCACACCCACCACTCCCGCGCACAGCATCGCTGGGCGACCTGG 534
Db 40081 GCGCACGACCGCTCGGCTCGCGCGCACCTCGCTGGCGCGCGCTGG 40134

RESULT 9
AX211739
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS Streptomycetes noursei.
SOURCE Streptomycetes noursei.
ORGANISM Streptomycetes noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source
1. .125401
/organism="Streptomycetes noursei"
/db_xref="taxon:1971"
/notes="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN
Query Match 50.0%; Score 266.8; DB 6; Length 125401;
Best Local Similarity 68.7%; Pred. No. 5.1e-25;
Matches 367; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
Qy 1 GGAACCACTCATCACCAGCGGAAACCGCGCGCTCGCCACCACTCACCACCACTC 60
Db 40009 GGCACCGTCTCTCACCAGCGGACACCGCGCGCTCGCGCGCTGCTCGCGCGCACTG 40068
Qy 61 ACCACCAACCAACCAACCACTCTCTCTCACCAGCGGAAACCGCGCGCGCAACCCCC 120

Db 40069 GTACCGAGCAGCGCGCGCCACCTGCTGCTGGCGCGCGCGCGCGCGCGCC 40128
Qy 121 CACGACACACCTCACCACCCCACTCCAAACAAAGGATCCACCTCACCATCACCACC 180
Db 40129 GCGCGCGCGCACTCCACCGCGCACTGACCGCGCTGGCGCGCGAGGTACCGTGGCGGC 40188
Qy 181 TCGACACACGAGCAACCCAGACCACTCCAAACAACTCTCAACACCACTCCCGCCCAACAC 240
Db 40189 TGGAGCTGCGCGACCGACCGCGCTCGCGCGCTGCTCGCCACCGTGGCGCGAACAC 40248
Qy 241 CCGCTCACCACCGTATCAGACCGGAGGATCTCGAGAGCGCAACCTCACCACCTC 300
Db 40249 CCGCTCACCAGCTGCTGACACCGCGCGCTCTGAGACGCGCAACCTCACCAGCTG 40308
Qy 301 ACCCGCACCACTCAACAGCTCTCGCGCGCAAGCCACAGCGCGCCACCTCTCCAC 360
Db 40309 AACCGGACCGCTCGCCACCGTCTACGCGCCCAAGTGGAGCGCGCTGGCACTTGCAC 40368
Qy 361 CAACTCACCACCAACACCCCGCTCACCAGCTTGTCTCTACTCTCTCGCGCGCGCCACC 420
Db 40369 GACCTCACCAGCGACCTCGACCTGGCGCGTTCGTCTCTACTCTCCACCGCGCGCTC 40428
Qy 421 TTGGCGGACCGCGCAAGCACTAGCGCGGAGCAAGCGCTACCTCAGCGCTCGCC 480
Db 40429 ATGGCGGACCGCGCGCGCACTACGCGCGCGCAACACCTTCTCGACGCGCTCGCC 40488
Qy 481 CACCACCGCACACCCACCACTCCCGCGCACAGCATCGCTGGGCGACCTGG 534
Db 40489 GCGCACGACCGCTCGGCTCGCGCGCACCTCGCTGGCGCGCGCTGG 40542

RESULT 10
AX211705
LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS Streptomycetes noursei.
SOURCE Streptomycetes noursei.
ORGANISM Streptomycetes noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 65140)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
Location/Qualifiers
1. .65140
/organism="Streptomycetes noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 8270 a 25171 c 22273 g 9426 t
ORIGIN
Query Match 48.8%; Score 260.4; DB 6; Length 65140;
Best Local Similarity 68.0%; Pred. No. 3.7e-24;
Matches 363; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
Qy 1 GGAACCACTCATCACCAGCGGAAACCGCGCGCTCGCCACCACTCACCACCACTC 60
Db 20690 GGCACCGTCTCTCACCAGCGGACACCGCGCGCTCGCGCGCTGCTCGCGCGCACTG 20749
Qy 61 ACCACCAACCAACCAACCACTCTCTCTCACCAGCGGAAACCGCGCGCGCAACCCCC 120

WRLGDEVADVRLPDGSRITGIGFGLHPALLEALAAASCLAGYSSPBGAVWISTDK
DVLRLTAGDRURVRLAPADGALANVOLADGRAPVAVSVVACRAIAVGEVSNALAR
NRDSLPHVDWA VEPVAPASPTERWISLIGADEPGARRYPDVASLARAVAGVSDVALV
CTETGDTGDRADVADVTRREALVQDMLEERLAGRLVTRTGQVTRAGDVEDVTL
GAAAVGLLRSAQSENPRGIVLVDVDEEASARILPAPVVASGEPOSAVDGQGVFVRL
GRVPAPAGAAATGANDRGTVLITGTSGLGFARHLVVEHGVNLLLSLRRGDAE
GAALEALITAGARVETAAADANDRAALAAVLADIPAEYPLTGVLTAGVLDDGLIG
SQTQRLSALVLPKVDATWNHETRLMDLSAFVLFSSIAAVVGGPGOSTYSAANHFL
DALAHRSRGILAAATSLAWGLSQTSGMTGHLDEADLKRARSFRPVESEOGPALLD
TALTVRAGLAVPMDLTALREOVQVPLLSKLARTPVRTAONAEVSESIAERLAG
LQDAQREAVLIGVQAAVVLGHSDATGIDRQSFALGDFSLTSVELRNQLGSTG
LRLPPLTVDPHTPVVALAEHLGEVRAELSDGASRPGVDFAAEVRADDDVRPADEV
ITTAADPKELILTAGSGFLGAFLLRLDMRTTARVHCLVRGDAEAERLKANAEWY
ITTAADPKELILTAGSGFLGAFLLRLDMRTTARVHCLVRGDAEAERLKANAEWY
SVNDDIDPERSVGLDLAEAPRLGLDEESFDALARTDVVYHNGARVHWLLPYETLKA
SNVGTTEVRLAARHRTVPVHVSTGVDFGVBEPVPLRVDTDPGAEISLPQSYLR
SKWAEQVTEVADRGLPVSVRVDSIGDRVNGACOTDFVWLTLKGLQIARSPKVG
TGRFHLLPVDVYSAAITGIRQPTGRTFHLFQNSLALSQCVELLRSLRGLYELDEV
DMDTQVLLVSGDNALLPLLDAPFEMMTSDTDGFPYPPIDTAAETVAALGEGIGIPVLT
ELFKEYVAFVVEEGHPPAAG"
complement (4732..4935)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00550 pp-binding,
phosphopantetheine attachment site, score 68.50, E-value
2.5e-17"
misc_feature
complement (4810..4857)
/gene="SC2C4.03c"
/note="P500012 Phosphopantetheine attachment site"
complement (5248..5268)
/gene="SC2C4.03c"
/note="P500340 Growth factor and cytokines receptors
family signature 2"
complement (7342..8286)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00698 Acyl_transf, Acyl
transferase domain, score 248.50, E-value 9.5e-71"
misc_feature
complement (8596..9867)
/gene="SC2C4.03c"

Query Match 50.0%; Score 266.8; DB 1; Length 24225;
Best Local Similarity 68.7%; Pred. No. 7.1e-25;
Matches 367; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 1 GGAACACCTCATCACCGCGGAACCGCGCCTCGCCACCCACCTCACCCACACCTC 60
DB 16508 GGCACCGTCTCATCACCGCGGACCGCGCCTCGGCACCGTCTCGCCGCCACCTC 16449
QY 61 ACCACCCACCAACCCACCAACCTCTCTCTCACCGCGGAACCGCGCCACACCCCC 120
DB 16448 GTCACCCAGACGGCAGCCGCCACCTTCTCTCTCACCGCGGACCAACGCCCC 16389
QY 121 CAGGCACAACCTCACCCACCACTCCAAACAAAAGGATCCACCTCACCATCACCC 180
DB 16388 GGGGCAACAGACTCAGATGACTACCGGCCTCGGGCCACCTGACCATCACCGCC 16329
QY 181 TGGACACCAAGCAACCCAGCAACCTCCAAACAACTCTCTCAACACCATCCCCCAACAC 240
DB 16328 TGGACACCGCGGACCGGACCGGCTTCTCGGTACTCGACGACATCCCGCGGAAC 16269
QY 241 CCCCTCACCACTGATCCACACCGGAGGATCTCGACGAGCCACCTCACCAACCTC 300
DB 16268 CCCCTCACCGCGGTCTGTCACACCGCGGCTCTCTCGACGAGCGGTCTCTCTCGCTC 16209
QY 301 ACCCCACCACTCAACAACTGCTCTCGCGGCAAGCCACACGCGCCACCTCTCTCCAC 360
DB 16208 ACCGAGGAGGCTCGCCGATGCTGCTCGGAGGTGACGCGGCGTGGAACTGAC 16149
QY 361 CAACTACCAACACACCCCTCTCACCGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 420
DB 16148 GAACTGACCAAGACCAACGACCTGGAGGATCTGCTCTCTCTCTCTCTCTCTCTCTCT 16089
QY 421 TTCGGCGCACCGGCAAGCAACTACCGCGGAGCAACGCTTCTCTCTCTCTCTCTCTCTCT 480
DB 16088 ATCGGCAACCGCGCGGCGGCAACTACGCGCGGCAACACCTTCTCTCTCTCTCTCTCT 16029

QY 481 CACCACCGCCACACCCACCTCCCGCCACACGACATCGCTCGGGCACCTGG 534
DB 16028 CAGCACCAGCGGACCTCGGACTCCCGCCCGCTCGCTCGGGCTGTGG 15975

RESULT 8
AF263912 123580 bp DNA linear BCT 24-MAY-2000
LOCUS Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
complete sequence.
AF263912
AF263912.1 GI:8050835
VERSION
KEYWORDS
SOURCE
ORGANISM
Streptomyces noursei.
Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 123580)
Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strøm,A.R.,
Valla,S. and Zotchev,S.B.
Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
Chem. Biol. 7 (6), 395-403 (2000)
10873841
20334850
2 (bases 1 to 123580)
Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strøm,A.R.,
Valla,S. and Zotchev,S.B.
Direct Submission
Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source
1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement (46..783)
/gene="nysP"
complement (46..783)
/gene="nysP"
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=11
/product="NysP"
/protein_id="AAF71762.1"
/db_xref="GI:8050836"
translation="MIELLPATVATEAAYDDRPDRLLSREVIARAVESQRE
FTTVHLLARRLRGLGHDRAILNRRGAPQPGIVGSMTHCAGYRAAVSPAEISA
AVSDAEPNGPLPAGLVNAILPSERPHLVAAHRPDVHMDRLLFSAKESVFKAYIP
LTQRELDSEASIVIDPTQGAFTARLLVPGPLGGRVTVFPGRHHSTPALTTAVHL
PAPTRRDRERHRLTVNSPLPRPTFG"
complement (867..2684)
/gene="nysG"
complement (867..2684)
/gene="nysG"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AAF71763.1"
/db_xref="GI:8050837"
translation="MASPDDLEERTAPRRRLVLLRPHRRSVVALVSMGVGGIVL
NAGPLLRGTVTLADVGLGVPAGPAGIDFAAIGRLVLLVLLVYVASLPMQAQR
LVASAVMTIHELREDADEKLTPLRPHRDQPGAGELLRTNDLQOTLQOTLAE
LITSFSLTMLVLMVLSPLAVVLLSVPSALIAARISKAQPHVAQHSANGTL
NAHVEVCTGHALIKGFORRAAEERFACNDADVTRAAKAQFASGAEPPVMFVANL
GYLVAVAVIGAWKVTGLTLDVQVAFILYARQFSQPIVEIASVAGRLSGIASARVF

REFERENCE AUTHORS JOURNAL REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 24225)
Seeger,K.J. and Harris,D.
Unpublished
3 (bases 1 to 24225)
Cerdano,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
Direct Submission
Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gi:12274798.

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the Frameplot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/>

<http://www.nih.gov/jp/>
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, ggg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 2c4.

FEATURES

source

1. .24225
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"
complement(2..1477)
/note="previously sequenced DNA fragment. EMBL:AF113603
Streptomyces coelicolor putative acyl-CoA carboxylase
complex A subunit (accA1) gene, complete cds."

misc_feature

misc_feature

2. .119
/note="nominal overlap with Streptomyces coelicolor cosmid
SCAH10"
complement(374..1471)
/note="pfam match to entry PF00289 CPase_L_chain,
Carbamoyl-phosphate synthase (CPase), score 585.30,
E-value: 3.8e-172"
complement(602..625)
/note="PS00867 Carbamoyl-phosphate synthase subdomain
signature 2"
signature 2
complement(1485..1489)
1774..1778
1786..1788
/gene="SC2C4.02"
/note="SC06272"
1786..1788
/gene="SC2C4.02"
/note="SC2C4.02", possible secreted FAD-binding protein,
len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222)
Streptomyces peucetius casius DnrW, 485 aa; fasta scores:

misc_feature

RBS

gene

CDS

opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa
overlap. Contains pfam match to entry PF01565
FAD binding 4, FAD binding domain. Also contains possible
N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted FAD-binding protein"
/protein_id="CAC22143.1"
/db_xref="GI:12274800"
/db_xref="SPTREMBL:Q9EX55"
/translation="MTEVSRRLKMGAAVSGALALPALGAPPATAPAAAGEDLPGP
AAAAARRPAPAPVHTVRSRDYQGLISGNSWGNPDYIRVVSAGVQVAVQAV
DKGLKIAVSGCHDEDFVANDVRVIMDMGMSDYVDREARFAVGARGLGTVYR
TLYKRGVVLPGCTPVTGAGHITGGYGLSRGLTVLHLYAVVHVHVGARGR
KVATREDDNPNRLWHAHTGAGCNFGVITRYLWLRSPDVGDDSTLLPAAPSEVLL
SDVSWDDLDDEASFTLLRNFTWHARNAPSDFGRFLFSQKTMHKAAGVPRMSTQ
VDAAPDADLDLDDYLAALSEGTVYHVGRYRPAWLYAVTENGSGFVTSVPRWKS
SAYRVNMPESQLRAVYRQLTDDYPGPGMIAIVFGGKINEVAGPDTATADRSIA
KMLYGLSDPADDALHQRIWEAREAYVASTGGVPRFGVNDGCIYINADADLADPA
LNRSIGPWHELYFKGNYPRQVKATMDPRNVFSHRLGIRK"
1948..2535
/gene="SC2C4.02"
/note="pfam match to entry PF01565 FAD binding_4, FAD
binding domain, score 6.50, E-value 1.1e-06"
complement(3514..9972)
/gene="SC2C4.03c"
/note="SC06273"
complement(3514..9972)
/gene="SC2C4.03c"
/note="SC2C4.03c", probable type I polyketide synthase,
len: 2152 aa; N-terminal domain almost identical to
TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2)
type I polyketide synthase (fragment), 720 aa; fasta
scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity
in 721 aa overlap and C-terminal domain similar to
SW:HEHTM ANASP (EMBL:L22883) Anabaena sp. polyketide
synthase HetM, 506 aa; blastp scores: Score= 761 (267.9
bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identity= 184/491 (37%), Positives= 278/491 (56%). Also highly
similar, in this same cosmid, to the N-terminal region of
SC2C4.04c and to SC2C4.05c. Contains pfam matches to
entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase,
PF00698 Acyl transf, Acyl transferase domain and PF00550
pp-binding, Phosphopantetheine attachment site and matches
to Prosite entries PS00606 Beta-ketoacyl synthases active
site, PS00340 Growth factor and cytokines receptors family
signature 2 and PS00012 Phosphopantetheine attachment
site"
/codon_start=1
/transl_table=11
/product="putative type I polyketide synthase"
/protein_id="CAC22144.1"
/db_xref="GI:12274801"
/db_xref="SPTREMBL:Q9EX54"
/translation="MSNEQKMYEYLKWVTTDLQKARQLAEAGNDFPVAIVGMACR
FPGVASDDLLRWLMDADGIDTDPTRGMDIDAIYDPERGPKGKSVTRKGGFLDGA
AOPDGFEGISPREALAMDPOORVLIETAWETFEQAGIDPATLRSOVSQVGVAVGVEQS
YLGLEGPEELGYLMTSKLSVASGRIATYTFEGPQVSVDTACSSSLVALHLAVQSV
RKSGSTWALAGSTVSGNFGVDFSRQGLAEDGRKIKSFAAADATAGASVGGULLY
EKLSDARRNRHVRVAVRGSVAVNQDASNGLTAPNPGSQERVIHQALAGALINASDVO
VVEAHGTCTRLGIDPIEAQALLSYGRQRPADRLPYLGLSKNSIKNIHTVAAAGGVYKIM
IMAMRHGILPKTLAVDPTMVDWEAGAVELLTARPWETCAPRAGVAFVSGSTN
AHVILERPAREPEEPADTAVLPLVLPWALSARTAPARQALRDLSHVREPRADP
LDVAYSLAERSALEHRAVVVGSDBRELLTGLRALAEGRTTPTGAVQTRQTKTAPL
FTGGQARTGMGLYRTPPAYAHAFDTITAQDLPHLDQPLHHTITTHGLHHTGNTQ
PALFATEVALYRLLETWGITPDYLAGHSIGELTAHSTGLITLQDACLTVTARALMQ
NLPTGTMIALQATEQEILPHLTGHEHLTAAINSPTSLVSGNTAQHATQLOA
QGRKTKLTWSHAFQMGMLHDHHTATQLTVEPTIPIVSLTGNATHTNLDRT
PTYWDOLRNTVRYTQALHTLTAGTYYTTEIGPDATLTPLTQOTLTDTTAPLRLHN
QPETHLLTALSLRNKGIPIWKNKLTFTTTPAKHTPLTYAFQHRFWFVTSVGAADA
SDLGESAGHPMLGAAVAVAGQVLFTRSLRSLTHPLMADHVDQTLVATATALVEM
AIRAGDGLSTVLDNLSDYPLVLPFGKGVHVQSVGSPGSLRSLYSIRSPDNADV
PWTNLNRGLSFRGPDAPFSLAQWPPAGAEVPLDQAYDKLAVQGVYVYGVYRGLGHEV

misc_feature

gene

CDS

SSTPGALPAAPMWHIDWPQDSTSLVTRRYAVVGADADDEAVLRAFLPDHATTAHA
 DLALRAAPDAVPTPGPALVVALIPGSPPAPEATSPDPVRAATLHGLSLVALIE
 DERLTGTTTAVTGAVGATOTDVPDLDLAGALMGLVRSASQSEYRFLRINDVDSPE
 SRAALATALASDEPQLALRAGWMLPALTPAERAGSPAPADAAPTANGMDGYLLTGC
 TADGRARVAVSHHARHGVRMLLVSRGPNPAERELTELVRKAGADAVRHLHEATL
 ADVKAVAGVPSHDPLTVGVVHTAGVLDGALARTLPERIDTVLRKAGADAVRHLHEATL
 GHPLRAFLVLSAAGLLGRPGQASVAAANALVDAFARLRAGLPAWSLAWGLMDERE
 GMAGGLDATALRRLRDGLLPMQADALLELDQALTYRDGPALLVPLVLLDSTALRRT
 ACENGPASVPLLRSLVPSRPHRTTATSVRDESGRGPASRLAALPAKERMVTLDD
 TVSEQVAGWLNELPAAIPDPERPHEIIGFDSLALTELRNLRLGLVGLRSLPSTLAFDHP
 TPKQVAGWLNELPAAIPDPERPHEIIGFDSLALTELRNLRLGLVGLRSLPSTLAFDHP
 LUTALDRPASGADSTRAAATTVADELDEATDEIIFAFLEQLEQEPDEPLGAPLDERL
 DEQP"

gene 13693..19212
 /gene="nida2"
 CDS 13693..19212
 /gene="nida2"
 /note="orf2"
 /codon_start=1
 /transl_table=1
 /product="polyketide synthase module 3"
 /protein_id="AAC46025.1"
 /db_xref="GI:2558839"
 /translation="MTAENDKIRSYLKRATBELHRTKERLAELESARHEPVAVGMAC
 RYPGVATPDDILWELVASGDTAISTFTDRGDLDGLYDPPSTFKSYVRHGGFLHD
 AAQFOAEFGIIPREBAMDPQORLLLETSEALEHAGIAPNLTARTTGTGVTGYMD
 DYGRCFPSLPVEGVVNGSVAGSAGRVATLGLQGPALVDTACSSSLVALHVLV
 QSLRGECDLALAGSTWMTPTVFVFSRQGLAPDCRCKSPADTAGTAWAEGVGV
 LIVERLSABRGLHRLVAVRGTAVNQDGSNGLTAPSGPAQQVIRKALSDAGLSAD
 DIDAVEAHGTGPDPIBAGALLATYGHPEQTPTWLGSLKSNIGHTQAAAAGIIGII
 KMIQAMHGTPLRLTHVDPKSKWKAGSLQLLAQAPWADPDRPRAAASAFGVS
 GTNAHAIIEBPTSLTEAPVDETPVPAHTPPVIAWPLSAHTPTALRAQAAARLWHLH
 TOADPODIAHALATYTRHFKRAVITGRTRTELTKLHTLDAIQTGAPHPRLTLFT
 GQAGHPGQSLYTTDPHFAALDEVCSELQRCGTQNLREVMFTPDQDPLDLTEYT
 QPALFALQTLRYTLTARGTQALHVGHSGEITAAHIAAGVLDLPDAARLITARAHLM
 GYUPLHGMKSVQAAHEHDLQALHTHGEIVNANGPTHCVLSPGTALBETAAHLREQ
 NVRLTWLVSHAFSALMDPLMGAFDRLTNLYNQPTPLISLNTGTALBETAAHLREQ
 PPETLTATALTHTGHTTTLHTTSQSHLIDLPYTFQDRYMAVPPRAAVGDLA
 DVIIDHARTVAFADAVQTAHQTTTLEIGHPHTLTLTLHHTLDNPTTITLHREH
 AVGTAVGHPLAANTEVETITVTWGMLSARALPWADHLVWDRGVAGTAVLEIV
 LEIGRIGFCIGELVLEAPLSWGDAPAVRTVVGALSPGTEVRSVLSYRDESBAM
 THAGSRARAGSEMGVFAELSGVPPPGAEVDISGVEYELFAAGVGYGGAFRG
 LAMVRRREDEIFAEVRLPDHADAAGYGLHPALLDAALHPITALDPLDGGHGLLPF
 VMTDLVRNCGNGHTLRVTRAQDPAQDQDDPGGAVSLAAADRQDGVFVARSILV
 RVSATQLPAAFPVPLHLDWPIIGFTSTGAGGDSRSLDLTELADVHLTVREVL
 EAVQHLREPVGGGLVLTREAGPEGGAPAGAAWGLVRSASQSHPEGRFVLID
 VGTBESSRALPAAVMAVDVAPQAVRQGRPSVPLVPLTADTRTEAPPGPGSAGL
 LQPDGTVLITGTGALAAVTHRLVARHGRLHLLVSRGTDAPQAAELASBELGALG
 DVEFRACDVTDRAAVQRLLAGI PAEHPLS CVVHTAGVLDGVI SAQTPERVAALVGP
 ADGALHLDLTDRLDRSVPVLVFSVSGVLSTGHAGYAAANAFDALARARRADGHP
 ALSLGMWMTGTMAAGLDRADTARTIRRSGLAPLDAETAMDILFDRLARPEPALPVR
 LDHAALEARARATLPEVRLDLAGVPSRAHPAPSGEGTATSTGTSGSLAERLARPTA
 ERPLLLELLEVTVAALVHGDPGAVGABRSFQAGFDSLTAVDLRLNRLNARTGLRLP
 ATLFDHPPTLALALLAGLDVPEPAADGPEDLAAAFDRLARGLAALDDRAARVGAEE
 LRSLTALTADPATGTGAQAGPEADGALTVEVTDRLRSASDDELFDLYDSEFR"
 19269..30233
 /gene="nida3"
 19269..30233
 /note="orf3"
 /codon_start=1
 /transl_table=1
 /product="polyketide synthase modules 4 and 5"
 /protein_id="AAC46026.1"
 /db_xref="GI:2558840"
 /translation="MSATNEEKLREYLRAMGLDHSARERLAELESARHEPVAVVGMA
 CRYPGVATPDDILWELVASGDTAISTFTDRGDLDGLYDPPSTFKSYVRHGGFLH
 DAAQFAEFPGISPREATAMDPOORLLLETSEALEHAGIAPNLTARTTGTGVTGMH
 HDYSHQVGSADASQGLGTAGSVASGRVATLGLQGPALVDTACSSSLVALHVA
 VLSRGECDLALAGSTWMTPTVFVFSRQGLAPDCRCKSPADTAGTAWAEGVGV
 MLVLSRDAERLGHRLVAVRGTAVNQDGSNGLTAPSGPAQQVIRKALSDAGLSAD
 DIDAVEAHGTGPDPIBAGALLATYGHPEQTPTWLGSLKSNIGHTQAAAAGIIGII
 KMIQAMHGTPLRLTHVDPKSKWKAGSLQLLAQAPWADPDRPRAAASAFGVS
 GTNAHAIIEBPTSLTEAPVDETPVPAHTPPVIAWPLSAHTPTALRAQAAARLWHLH
 TOADPODIAHALATYTRHFKRAVITGRTRTELTKLHTLDAIQTGAPHPRLTLFT
 GQAGHPGQSLYTTDPHFAALDEVCSELQRCGTQNLREVMFTPDQDPLDLTEYT
 QPALFALQTLRYTLTARGTQALHVGHSGEITAAHIAAGVLDLPDAARLITARAHLM
 GYUPLHGMKSVQAAHEHDLQALHTHGEIVNANGPTHCVLSPGTALBETAAHLREQ
 NVRLTWLVSHAFSALMDPLMGAFDRLTNLYNQPTPLISLNTGTALBETAAHLREQ
 PPETLTATALTHTGHTTTLHTTSQSHLIDLPYTFQDRYMAVPPRAAVGDLA
 DVIIDHARTVAFADAVQTAHQTTTLEIGHPHTLTLTLHHTLDNPTTITLHREH
 AVGTAVGHPLAANTEVETITVTWGMLSARALPWADHLVWDRGVAGTAVLEIV
 LEIGRIGFCIGELVLEAPLSWGDAPAVRTVVGALSPGTEVRSVLSYRDESBAM
 THAGSRARAGSEMGVFAELSGVPPPGAEVDISGVEYELFAAGVGYGGAFRG
 LAMVRRREDEIFAEVRLPDHADAAGYGLHPALLDAALHPITALDPLDGGHGLLPF
 VMTDLVRNCGNGHTLRVTRAQDPAQDQDDPGGAVSLAAADRQDGVFVARSILV
 RVSATQLPAAFPVPLHLDWPIIGFTSTGAGGDSRSLDLTELADVHLTVREVL
 EAVQHLREPVGGGLVLTREAGPEGGAPAGAAWGLVRSASQSHPEGRFVLID
 VGTBESSRALPAAVMAVDVAPQAVRQGRPSVPLVPLTADTRTEAPPGPGSAGL
 LQPDGTVLITGTGALAAVTHRLVARHGRLHLLVSRGTDAPQAAELASBELGALG
 DVEFRACDVTDRAAVQRLLAGI PAEHPLS CVVHTAGVLDGVI SAQTPERVAALVGP
 ADGALHLDLTDRLDRSVPVLVFSVSGVLSTGHAGYAAANAFDALARARRADGHP
 ALSLGMWMTGTMAAGLDRADTARTIRRSGLAPLDAETAMDILFDRLARPEPALPVR
 LDHAALEARARATLPEVRLDLAGVPSRAHPAPSGEGTATSTGTSGSLAERLARPTA
 ERPLLLELLEVTVAALVHGDPGAVGABRSFQAGFDSLTAVDLRLNRLNARTGLRLP
 ATLFDHPPTLALALLAGLDVPEPAADGPEDLAAAFDRLARGLAALDDRAARVGAEE
 LRSLTALTADPATGTGAQAGPEADGALTVEVTDRLRSASDDELFDLYDSEFR"
 19269..30233
 /gene="nida3"
 19269..30233
 /note="orf3"
 /codon_start=1
 /transl_table=1
 /product="polyketide synthase modules 4 and 5"
 /protein_id="AAC46026.1"
 /db_xref="GI:2558840"
 /translation="MSATNEEKLREYLRAMGLDHSARERLAELESARHEPVAVVGMA
 CRYPGVATPDDILWELVASGDTAISTFTDRGDLDGLYDPPSTFKSYVRHGGFLH
 DAAQFAEFPGISPREATAMDPOORLLLETSEALEHAGIAPNLTARTTGTGVTGMH
 HDYSHQVGSADASQGLGTAGSVASGRVATLGLQGPALVDTACSSSLVALHVA
 VLSRGECDLALAGSTWMTPTVFVFSRQGLAPDCRCKSPADTAGTAWAEGVGV
 MLVLSRDAERLGHRLVAVRGTAVNQDGSNGLTAPSGPAQQVIRKALSDAGLSAD
 DIDAVEAHGTGPDPIBAGALLATYGHPEQTPTWLGSLKSNIGHTQAAAAGIIGII
 KMIQAMHGTPLRLTHVDPKSKWKAGSLQLLAQAPWADPDRPRAAASAFGVS
 GTNAHAIIEBPTSLTEAPVDETPVPAHTPPVIAWPLSAHTPTALRAQAAARLWHLH
 TOADPODIAHALATYTRHFKRAVITGRTRTELTKLHTLDAIQTGAPHPRLTLFT
 GQAGHPGQSLYTTDPHFAALDEVCSELQRCGTQNLREVMFTPDQDPLDLTEYT
 QPALFALQTLRYTLTARGTQALHVGHSGEITAAHIAAGVLDLPDAARLITARAHLM
 GYUPLHGMKSVQAAHEHDLQALHTHGEIVNANGPTHCVLSPGTALBETAAHLREQ
 NVRLTWLVSHAFSALMDPLMGAFDRLTNLYNQPTPLISLNTGTALBETAAHLREQ
 PPETLTATALTHTGHTTTLHTTSQSHLIDLPYTFQDRYMAVPPRAAVGDLA
 DVIIDHARTVAFADAVQTAHQTTTLEIGHPHTLTLTLHHTLDNPTTITLHREH
 AVGTAVGHPLAANTEVETITVTWGMLSARALPWADHLVWDRGVAGTAVLEIV
 LEIGRIGFCIGELVLEAPLSWGDAPAVRTVVGALSPGTEVRSVLSYRDESBAM
 THAGSRARAGSEMGVFAELSGVPPPGAEVDISGVEYELFAAGVGYGGAFRG
 LAMVRRREDEIFAEVRLPDHADAAGYGLHPALLDAALHPITALDPLDGGHGLLPF
 VMTDLVRNCGNGHTLRVTRAQDPAQDQDDPGGAVSLAAADRQDGVFVARSILV
 RVSATQLPAAFPVPLHLDWPIIGFTSTGAGGDSRSLDLTELADVHLTVREVL
 EAVQHLREPVGGGLVLTREAGPEGGAPAGAAWGLVRSASQSHPEGRFVLID
 VGTBESSRALPAAVMAVDVAPQAVRQGRPSVPLVPLTADTRTEAPPGPGSAGL
 LQPDGTVLITGTGALAAVTHRLVARHGRLHLLVSRGTDAPQAAELASBELGALG
 DVEFRACDVTDRAAVQRLLAGI PAEHPLS CVVHTAGVLDGVI SAQTPERVAALVGP
 ADGALHLDLTDRLDRSVPVLVFSVSGVLSTGHAGYAAANAFDALARARRADGHP
 ALSLGMWMTGTMAAGLDRADTARTIRRSGLAPLDAETAMDILFDRLARPEPALPVR
 LDHAALEARARATLPEVRLDLAGVPSRAHPAPSGEGTATSTGTSGSLAERLARPTA
 ERPLLLELLEVTVAALVHGDPGAVGABRSFQAGFDSLTAVDLRLNRLNARTGLRLP
 ATLFDHPPTLALALLAGLDVPEPAADGPEDLAAAFDRLARGLAALDDRAARVGAEE
 LRSLTALTADPATGTGAQAGPEADGALTVEVTDRLRSASDDELFDLYDSEFR"

Query Match 53.5%; Score 285.8; DB 1; Length 41097;
 Best Local Similarity 73.4%; Pred. No. 2.8e-27;
 Matches 394; Conservative 0; Mismatches 137; Indels 6; Gaps 2;
 QY 1 GGAACCACTCTATCACCAGGCGGAAACCGGCGCTTCGACCCACCTCACCCACCTC 60
 DB 33786 GGCACCATCTCTATCACCAGGCGGCAACCGGCGCTTCGACCCACCTCACCC 33845
 QY 61 ACCACCAACCAACCAACCACTCTCTCACAGCGCAACCGGCGCTTCGACCCACCT 120
 DB 33846 G---CCCAAGGCGCAACCACTCTCTCACAGCGCGCAAGGCGCTTCGACCC 33902
 QY 121 CACGCACCAACCTCACCAACCACTCTCAACCAAAAGGCATCCCTCACCATCACCA 180
 DB 33903 GCGGCCACCACTCAACCACTCTCACCACTCTCACCACTCACCACTCACCA 33962
 QY 181 TCGCACACCAACCAACCACTCTCAACCACTCTCAACCACTCTCACCACTCACCA 240
 DB 33963 TCGGACCTCACGACCGCCCAACCACTCTCAACCACTCTCACCACTCACCA 34022
 QY 241 ---CCCTTACCACCGTCTATCCACACCGGCGGCTCTCGACGCGGCGGCGGCGG 297
 DB 34023 CCACCCCTTCAAGGCACTTTCACACCGCGGCAACCGGCGGCGGCGGCGGCGG 34082
 QY 298 CTCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
 DB 34083 CTCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 34142
 QY 358 CACCAACTCACCCCAACCACTCTCACGCGCTTCGCTCTCTCTCTCTCTCTCTCT 417
 DB 34143 CACCAACCAACCAACCACTCTCACGCGCTTCGCTCTCTCTCTCTCTCTCT 34202
 QY 418 ACCTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
 DB 34203 ACCTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 34262
 QY 478 GCGCACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 534
 DB 34263 GCGCACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 34319

RESULT 7
 SC2C4/c 24225 bp DNA linear BCT 12-MAY-2002
 LOCUS Streptomyces coelicolor cosmid 2C4.
 DEFINITION Streptomyces coelicolor cosmid 2C4.
 ACCESSION AL512902 AL645882
 VERSION AL512902.2 GI:20520683
 KEYWORDS accA1; secreted FAD-binding protein; type I polyketide synthase.
 SOURCE Streptomyces coelicolor A3(2).
 ORGANISM Streptomyces coelicolor A3(2).
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 1 to 24225)
 AUTHORS Kinsahi, H., and Hopwood, D.A.
 TITLE A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
 JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
 MEDLINE 97000351
 PUBMED 8843436

```
QY 121 CAGCGACACACCTCTACCCACCCAACTCCAAACAAAAGGATCCACCTCACCATCACCAC 180
    |||
Db 96548 GCGCGCACCGAACTCGTAGCGGAACCTCACCGCAGCGCGCGCCACGTCACCGGC 96489
    |||
QY 181 TCGGACACCGAGCAACCGACCACTCCCAACACTCCTCAACACCACTCCCGCCACACAC 240
    |||
Db 96488 TGGGACACCAACCGACCGCGGAACTCGGCACCTCTCACCACCACTCCCGCCACGAAAC 96429
    |||
QY 241 CCCTCACCACCGCTATCCACACCGCAGGATCCTCGAGAGCGCCACCTCCACCAACCTC 300
    |||
Db 96428 CCCTCACCACCGCTATCCACACCGCAGGACCTCGAGAGCGCCACCGAGAGCTG 96369
    |||
QY 301 ACCCCACCACTCAACAACTCTCGCGCGCAAGCCACGCGCCACCTCTCCAC 360
    |||
Db 96368 ACCGACACCACTCCACACCGCTCTCCAAACCAAGATCGACGACGACCCACCTCCAC 96309
    |||
QY 361 CAACTACCCACACACCGCCCTCAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
    |||
Db 96308 CAACTCACCCTCCACACCGCCCTCAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 96249
    |||
QY 421 TTCGGCGCACCGCGCAAGCACTACGCGCGAGCAAGCGCTTCTCTCTCTCTCTCTCTCT 480
    |||
Db 96248 CTCGGACCGCAGGACAGCGCACTACGCGCGCGCCACACCTACTCTCAGCGCCCTCG 96189
    |||
QY 481 CACCACCGCACACCCACCTCTCCCGCCACGAGATCGCTCGGCGACCTGG 534
    |||
Db 96188 CACCACCGCGCACCCACGAGACACCCCGCACCTCTCTCTCTCTCTCTCTCTCTCTGG 96135
    |||
```

```
RESULT 6
AF016585 41097 bp DNA linear BCT 07-DEC-1997
LOCUS Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)
DEFINITION gene, partial cds; polyketide synthase modules 1 through 7 (nidi)
GENES complete cds; and N-methyltransferase homolog gene, partial
cds.
ACCESSION AF016585
VERSION AF016585.1 GI:2558836
KEYWORDS Streptomyces caelestis.
SOURCE Streptomyces caelestis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 41097)
AUTHORS Kakavas, S.J., Katz, L. and Stassi, D.
TITLE Identification and characterization of the niddamycin polyketide
synthase genes from Streptomyces caelestis
JOURNAL J. Bacteriol. 179 (23), 7515-7522 (1997)
MEDLINE 98053867
PUBMED 9393718
REFERENCE 2 (bases 1 to 41097)
AUTHORS Kakavas, S. and Stassi, D.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park
Road, Abbott Park, IL 60064, USA
FEATURES
Source 1. 41097
/organism="Streptomyces caelestis"
/strain="NRRL-2821"
/db_xref="taxon:36816"
complement(<1..396)
/gene="nidi"
complement(<1..396)
/gene="nidi"
/note="ORF6; similar to tyll"
/codon_start=1
/transl_table=11
/product="cytochrome P-450 hydroxylase homolog"
/protein_id="AAC46023.1"
/db_xref="GI:2558837"
/translation="WVDSVTGPMELSKDANAKELLDWFSHNTHPWFVWDEGRQAWQV
FRYDYLTVSNHPFSSDFTEVPTPELEMLFGITGALDPPAHGPMKRLVSOAF
TPRRMAGQEQIRIVIAEELLDVRGQKITA"
```

gene
CDS

```
627. .13649
/gene="nidi1"
627. .13649
/gene="nidi1"
/note="orf1"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translation="WAGHGDATQAQDAEKSESDGSDAIAVIGMSCRFPGAPGTAEFW
QLSSGDAVVTADARRGTIDAPDAAFFGMSPREAAATDPOORLVLEUGBAL
EDAGTVPSLEGEAASVFGAMNDYATLHHRAGATPTDTYATGLQSHANRLSYFL
GLRGSLLVDTGSSSLVAVALVESLRGSGTIGALAGGNLVLAEEGSAEMRVAL
SPDGCHTFDARANGVVRGEGAI VVLKPLDALADGDRVYCVVRGVA TGNDDGGPGL
TVPRAGQEAIVRAACDOAGVPRPADVRVELHGTCTGAGDPVEAEALGAVYGTREAN
BPLLVGSVKTVNIGHLEAGAGIAGFVKAALC LHERALPASINFETNPALPLERLKV
QTAHAALQFGTGGPGLAGVAFGMGNCVHVLBETPGGRQPAETQADACLFSSAP
MLLSARSEQALRAQAARLREHLESDGAPLDIAYSLATTRTFRFAAAVPCGDPDL
SSALAAAGQTPRGVIGSTDADGRLALLFTGGAQHPGMGOELYTTDPHFAAHLV
VCELRQCTGNLRVEMFTPDQDLDRTEYTOPALFALOTALTALYTLTARGTQAHVL
GHSVGEITAAHTAGVLDLPDAARLITARAHWGOLPHGGMLSVQAAHDLQDLAHTH
GVETIAVNGPCHVLSGPRTALETAQHLREQNVRHTLWKVSHAFSALMDPMLGAFR
DTLNTANTOPPTPLISNLGQIADPNHLCCTPDYWDIARHTVRFADAVQTAHQSTT
TYLEIGHPTTLTLHHTLDNPTTPTLHREPEPETLTOAIAAVGRTDGDIAVLC
GASRPRVLPETYAFQRRTHWAPGLTPNHADPADRAAPQRAMAVGPVSREALVRLVG
ETTSVGLGDPEVALDRPPTSQGLDSMTAVELAGLLGTAAGVALDPTLVYELPTPR
AVADHLAKTLGESAADQOVNGRTGEAAKAGDPIAVIGICRPPGAVATPDLME
LVASGTDALSTPPTDRDMDLGLYDPFSTPGKSYVRHGGFLHDAAPQDAEFPGISPR
EATAMDPOORLLLETSEWALERAGVVVPSLRGGRTGVFVFTTAPYGRPLHSGTGYE
GFLTGTTSVASGR IAYALGTRGPALTVDTCSSSLVALHLAVQSLRGECDLALAG
GTVMSGPMFVFSRQRLAPDGRCKAFSADADGTAWAEGVGLLVERLSDAERLGH
RVLA VVRGTAVNODGASNLTPASGPAQOQVTRDALSADGASDIDIAVEAHGTGTAL
GDPIEAGKLLATYGHKQTPVWGLSKNSNIGHTQAAGIAGIIMVQALRHDTLPRT
LHADPSKLVMDQQLDTPADPWPADPDRPRAGISAGVGNATNAHVLEVEEPAP
EAPARTQAPPVIAWPLSAHTPTALRAQAARLDSMLHHTQADPDQDIAHALATTHF
KHRAVITGRTELKTLHDLAIOGTAPHPRLTLFTGQAQHRGMGOELYATDPH
FAALHVECELRQCTGNLRVEMFTPDQDLDRTEYTOPALFALOTALTALYTLARG
TOAHLVLSGVEITAAHAGVLDLPDAARLITARAHWGOLPHGGMLSVQAAHDL
DQLAHTHGVETIAVNGPCHVLSGPRTALETAQHLHQQGIRHTLWKVSHAFSALMD
PMLGAFRDTLNTOPPTPLISNLGQIADPNHLCCTPDYWDIARHTVRFADAVQ
AQHDQTTLYLEIGHPTTLTLHHTLDNPTTPTLHREPEPETLTOAIAAVGRTDGDIA
TLHTSTPQTHLDPPTYPFQDRYRMEPVRAVQSGQGDRLRYRVVWEAAPEDHT
VPSARWLLGSDPADAGILVGEOLVGHGAQVORLDIPAPACRRLVPSCELSVPR
GPSHTGFCGSLRSPRAEAVALLQALGDAGADAPLMIATGVAQAQAGEASVGGGA
QLWGLQVAGLEADRWCGLVDLPADPAALRGLRLVLSNAPQVAVIRASGVFR
VVPAPGRPVNRDWPAGTGLITGTTGALGQVARRLRLAGPHLLHAGRRNGSGAAE
LVDELTALEAETVAAACDARDALVSLATIPERHPLTAVLHGAQVLDLDDGLDGLTP
BRIDALRAKATAARHLDELTAIDLDAFVLPSSIVGVNGGQATYAANAALDALA
HGRARGQATSIAGPWAGSMAAGAKSLARDGIDVLEPEEALDLEQAVRDET
TLVADVDWETFLSSASRRFQLPLFGRI PAARA VAVSGDSSDDVPVWLSPOLSA
ERQLLDLVRTEAASVLRHGSATVDAETAFRAGGPDLSLTLVELNRLSATGLRLS
NTLLDPDTPAAVAYLEGEQLLGTPESSAAASHGDAGARTVVASDFVAVGWACRYPG
GVATPDHDELWELVASTDAISTPPTDRDMDLGLYDPFSTPGKSYVRHGGFLHDA
DAEFPGISPREATAMDPOORLLLETSEWALERAGIAPHTLRTTRGVFTGASQSQYQ
OSREAADKYGGHLLTGLIASSVMSGRVATLGLQGPALTVDTCSSSLVALHLAVQSLR
RGECDLALAGGSTVMSPTVFSRQRLAPDGRCKPDAOTADGTANAEVGVLLVE
RGLDRLGRHLVAVRGTAVNQDSNGLTAPSPAQQQVIRDALSDAGLSADDIDA
VEAHGTGTPGLDPIAEALATYGHPEPTVWGLSKNSNIGHTQAAGIAGIIMVQ
AMHGTGTPRLTHVDAPTSQDWEAGSLQLAEAREPAPDRPRAGISAGVSGTNA
HVLEEPAPAEAPARTQAPPVIAWPLSAHTPTALRAQAARLDSMLHHTQADPDQDIA
HALATCTCFKHRAVITGRTELKTLHDLAIOGTAPHPRLTLFTGQAQHRGMGO
GOELYTTPDPEAALDELCELRQCTGNLRVEMFTPDQDLDRTEYTOPALFALOT
ALYRLTHAGTQAHVLHLSGVEITAAHAGVLDLPDAARLITARAHWGOLPHGGM
LSVQAAHDLQDLAHTHGVETIAVNGPCHVLSGPRTALETAQHLREQNVRHTLWK
SHAFSALMDPMLGAFRDTLNTANTOPPTPLISNLGQIADPNHLCCTPDYWDIARH
TVRFADAVQTAHQSTTLYLEIGHPTTLTLHHTLDNPTTPTLHREPEPETLTOA
LATLHTTGHHTTTPHPSIPAOQVSLPVPFORVAMPNSAAHICRSDAEAATRLGLA
RQHPFLTGATSI VSGGATLLTGRVSLAAHPLADHTVAGAVLPFGTAFADLLHVAE
ETGSGTVELLTHVLPVLPAQAGVQLQVLEAPDERGRRATVAARPEETGAETWTH
AAGVLAGEAPAAPADVADWAGAAPPPGAEAVDVDELYAABFAAYDYGPAETGLTG
AWRREEVFDVRLPEAVAGSGEERFVHPALFADALHPWRVGGPQSGPDDAGG
QSTLLPFSWQGLTYGTGAEQVRVRLAPSDGAFSVRAADPAGHVPALDALVLRVP
```



```

/protein id="BAB69187.1"
/db_xref="GI:15823970"
/translation="MSQDRHDCGTAALGGPSHGGRLLAVSDIHVRHQRNDIVQDLHP
GCAEDWLVAGDVASVADITWALSLLRRDRPAKYIWAQGNHSLWTAPGDDDLRGK
YHLDVCLCGAVVTPEDYPVWAGAGGVAVAPLLDYDYSFLPPGAWSKSHALRLA
DDAGVITDEFGLVPPDPYPSRAWRARVATRLADLADLPVTVNHFPLVREPT
RYRHPFALWCCTTATADWHLRYRAAAVHGHLLHPTIRCDVPHHEVSLGYPREW
QYGRRPDPVQILTPAPAPAPVPVPVPPGPHGTASPADRRP"
gene
3213..3971
/gene="olmC"
/gene="olmC"
/gene="olmC"
/codon_start=1
/translation="thioesterase"
/protein id="BAB69188.1"
/db_xref="GI:15823971"
/translation="MSANRVGNLWVRQHPHAPTAAATKLVCILPHAGGSASYFLPVSKA
LAPGVDLAIQIPGQDRHKECINIAELADAVTAVLEPLADRLPLVFGHSMGATLA
FEVALRLERKGIYPLAFASGRRAHRSDESVHLRGDDDEIIAELKTLSTGTRSEVLGD
BEILRWLPAIRSDYKAETRYAGGARLTAPIHAHFGVDEPRVSLDEAQAADHTTG
HPEILRSHPCGHFYLNDQAPRVIEETARVSSVSAS"
4643..7438
/gene="olmRI"
/gene="olmRI"
/gene="olmRI"
/codon_start=1
/translation="luxR-family transcriptional regulatory protein"
/protein id="BAB69189.1"
/db_xref="GI:15823972"
/translation="WLIGRQBIIDRLRAAYSSVRSGRAVLVESAAGCKGSEILED
LVGRHGLIARATAIAEERTPLSVLRVNGAPPPDDIVRSEHTMSDRAAEMIS
TEGRDKARDAAYELMTFCTEVRALAQAPLVAVVDDIHGDDLSRLQLHLARHC
RNVPLITQSLGAPADRTFGAELLRQPNLEHILRPFQDDIADMLAEAGRADRG
QVTERYFVAGGNPLLLRALLEADLEPDDTDQDVGPRAGAYGRAVLAACLDRGG
EKAHEVAGALVGLGAFTHILLSRLIDI SAAALTENRALRAAGIVGTGTFQHPAAQA
LVLDSLEPHRLRHLORSAALHAGCAPAAALAEHLHEAGQADGAWAVGLVQOAAEQ
AVLDBARRVSYLELAHAASADPRGRTEIKILGSVTRVSAADAEHVVDDALRGFA
GLITPSDTGLAALLHGRLEQASVLEQGRPGDGDADTAARQLAEIRDSVLR
GAWCAGTRAHNPAGRSTAGTGLRAWLSGSHALLAQRIAGTEGLLEIITHLD
ATFDVLNVAWTLALGAVERALHWCDFTLREAKRDAPGWEAVRAVRAEGALRGNN
LAEANDERCAFLVYPRHRSALEGSVSLVTVQIARGKYDADAQRLSRPNWADSLPD
THWIDYLRARGVLYMATRYQHSALRDHEVGRLEAERMGTDWPAIIPWRTDTAEALLR
LGEROGLRLAQEOLAMVODGAPRIKGISLRLOGLASBPAAIRGLNRAVVELSPADD
RIELESRTPLDADARETROTQAAWVLRARQLAKSCGAQPRDRISPDRAAGSGE
LFGEPERNADHGLSESERRVAGLAASGYTNREISVRLIITWSTVEQHLTRYKLIK
INGRELLAEARAGVQV"
7658..10486
/gene="olmRII"
/gene="olmRII"
/gene="olmRII"
/codon_start=1
/translation="luxR-family transcriptional regulatory protein"
/protein id="BAB69190.1"
/db_xref="GI:15823973"
/translation="WLASAGGEGGVVLVAGPGVGRVTRTALLDAFLRKATESGARVCAA
AGSSATEAEFVGVLFPPTPLGNGFLKRAIGSRDGPGLDRLDRLDLCQSLRR
RPTVIAVDDAQADAAASLFLYLRRLRTPAIVMVLVITEPTGPDALPPSFQALLRHP
QCRRLLGLDPLSDRGTQVTVSAQVSGSDVRPLADFDHAI SGGNPLLVHALLADRAHR
LVSPVSPVTVPSVSSAASASASASASASASASASASASASASASASASASASASAS
DVACGFAVLGESATPARVACLVRGDADAVDVMTALDTAGLAGGTGFRSALVGVKALLE
HYDVTRELRERARALLHGDGAPVAVARHLLAPVAPFPWAGRTLLAAAEAKMEGR
EELSCLDLRQASATTEQERAAVAVARVKIGWEIDPRLVSPVDELVAALRRGDRS
GDAAMIVKHLWQDRFPEAVESLTLAAAGDAGEPGVEAEHLFVRHLWRHSCPPLEAS
VPEGNPTPTPRVLRVSPASVAGEVLGKLTDEGDDHTVAMAEIEILRGCRYSYTV
EELAEALLAVLTPVGRPDRAAFWCGLLQQAQTRSTGTSVAIISGTRAEALTROQLPA
VQOYARSALSAISLPQWGVAGVSPILAALVHSTTASGMSGVAGAWLNQAVPDGWFTRH
GLLYTHARGHYHLAVDRPAAALDDFLTCCGLAKEMGVPTPLPWRTSAAQAQLALGN
PQARALVREQLRGCGGTARAGVALRLUATGEPGRPALRESVNLLESQGDQVE
LURADLRQCLDHWGAPAKARMTARHARNVADACGGDALFRKLFDEVPDREGPDG
LTRGEQGSASLTAAERVVTAALAGHSNREIQRKLFITKSTVEQHLTRYKLVGRNR

```

```

ADLGLLAGTNLGAGTTPAMAMASSAVAG"
10933..11271
/codon_start=1
/translation="p450-like protein"
/protein id="BAB69191.1"
/db_xref="GI:15823974"
/translation="MEVGGOTIRAGEVLCSTLSTANRDEGAFTDDELDLRRDARSHL
SPFGIHQICIGSLARAEPQIAETLLRRLRLAASFEELRFRDGVGVGVGVKELPD
PCPALPPDRP"
11458..29898
/gene="olmAI"
/gene="olmAI"
/gene="olmAI"
/codon_start=1
/translation="modular polyketide synthase"
/protein id="BAB69192.1"
/db_xref="GI:15823975"
/translation="MACRLPGAAGPAEPLALLRNGEDAVGAAPGDRRYAPRRRGGFLDS
VDEPDAGFGISPREAAMDPOQLVLELSWEALEDAGI PPARLGGSGAGVFPVGAIAID
DYATLSRAAGADAATPQTTCINRGIIANRVSYTLGLRGPSTFTVDSGSSSLVAHLLA
AESLRGECSSVALGGVNLIAEDSTVAVERFGLSPDGYCITTFDARANGVVRGEGG
VVLKPLADAIADGDDTHCILAGSANNVGGEGLTVPDRAGQEAVALTAAYEQAGISF
AAGVVELHGTGTDPDEAAVAGVLAGRAGPAGQPLLVGSVKNTNI GHLEGAAGIAG
LLKAVLAVHREI PATLHASPNAIPMAELGIGVNTALTTPWEQOAGPLVGVSSFGM
GGTNCHVLTETPAGYDAAAPATAGIRPAGTPAPILLVSRDEQALRAOARLQDLDT
HPGLPLRVANTHAAARAPLARRAVLLAADREDVSAGLARLADGTGADLVRGTVSPG
ALAFVFTQSGORFGMTAELYRTSAEYAAALDEVCASIDPMLRRLRDLVLPAAEGSDE
AALLDRFTQFALFAVEALFRYAEHCGLTFRFLSHSGVGLAAHVAAGVLSLADAC
TLVAARGLMQDRPATGAVSQVQATDELMPLTFEPRRI AAGLTFGAPRIPIVSNETGL
AVAAHWAARGKTKRLRVSHAFPHMDMLTEFFRI AAGLTFGAPRIPIVSNETGL
LTEDARSPEYVGRHAAVFLDQVRLLEQVTTLELGDGPDLSAMARDCLRDAG
TVCPLLRGSRTEPEALLAALAGLVGVPMWERTATEAGARRVPLTYAFORRRH
WLPEPSASRPVNVNGHAGQAGHAGQAGHAGHAGHAGHAGHAGHAGHAGHAGHAGHAGH
DRITTELVDRATAQVLGHASPDVATHEAFRELFSSLMIAELRGLEALTRGRRVPTT
LLFHPITDALARLEADAEERTDAGPVSPAAAADDDPIVIMGGRCLRGVSPASPE
LWTFVAAAGDAISALPVDGRWEPAADPGEGGLADVAEDFAGFGISPRGALMDP
QORLLETISWEALPRAGVDALKRSRTGVPFVAGSPSEYPRLHETSADGHLVGTGA
PSVLSGRITAYVLGLEGPALTVDACSSSLVALHLLAAQALRGCECDLALAGGVTVWATS
GMFAFARQGLADPGRCKAFADAGDTGMEGVGMLAVQRUSDADVRDGRPVLAVRG
SAVNSDGSNGLTAPNGPSQQRVIRQALASAGLSAADVMVMEHGTGTSGLDPIEBAQ
LLATYQDRPDRPLGLWSKNSHTQAAAGVAGIKSVLARNGLPLKTLHVDPT
PEVWSAGABELLTGREGWPTDGRPRAGVAFGISTNAHVILQAPSVBEPVTPA
DDGLLVPMVVSAREEALRAOARLADHVRSLGAPADVGLSLAVTRAGLHRAVLVA
EDERAGFLHLLALAEGRTPAGVVEGTTAFLESGGSGORAGMRELYSPVPAAFD
EVCARLGLCALREVVGAGLDRTDTEAQPALFALEVALFVLSVSGVRADVLIGH
SIGELAAAYVAGVMSLADACRIVAARGLMQALPEGMAVAVEAEDLPDLPGGVSF
AAVNGPSRLVSGDEEPTVLAQAPAEQGRRTKQTVSHAFHARMEMPLAEFTLA
AVEFRPPRIPVSNVTGEMAGDEFTTPAYVVRHVEAVFADGVGAHAAGVRRYLEL
PGDGLTTLTADTADGRDVAAPALRGDOPETRTVTVAVGRVATGTLVDMVPAFA
PYGARPTLPTVAFQHRVWLEPVAPSPAPPANPLRYRTAMASVPDGRPSLSCTWAV
VVPAGSYEDIVAGVTAALRDHGRPVVVEVDPPAGRADVADTLAALAEADGAGVVS
LLGTAGSSUSSGATAVLVQAWTDLGGGARLWCLTRGAVSVSPSPDPAQAQALWGM
GRVAALEHPEWGGVLDLPVDDRAMTRCALVLAGGSGEDQAVRAAGLFPARRLGRA
GTRADAAPAGSPWRTDGTVLVTGCTGALGALHQAOWLAAGAAHVLLTSRRGPADGA
AELTARLETCTVVAACDVTDRDALAEELSGLPADRLTGTVHVAAGVLDVGLDLSL
TPDRPAAVAGPKVTVGAHMLHETLDRDLDSAFVLSFSSPAGTVGLAPQANYAANVLDL
LAVRPAAGLGPATAVAMGWSGAAADTEARQLARTGLVPLAGDPAALAAALGQVMD
DGETAVTVADVDWDRFAAGFGPRSPPLLTGVPEVRRTEERTETGAGAPAEETDRUGGL

```

Query Match 62.8%; Score 335.6; DB 1; Length 104326;
 Best Local Similarity 76.8%; Pred. No. 1.5e-33;
 Matches 410; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 1 GGAAACCACTCATACCGCGGGAAACCGCGCCCTTCGCCACCCACCTTCAACCCACCACTC 60
 |||||
 Db 96668 GGCACCGCCCTCATCACCGCGGACCGCGACCCCTCGCGGCCCTCTCTCGCCGCCACCTC 96609

Qy 61 ACCACCAACCAACCAACCAACCACTCTCTCACACCGGACCGGCCCCCCCCACACCCCC 120
 |||||
 Db 96608 GTCAACCGGCACCGGGGTGACCAAGCCTGCTGCTCACACCGGACCGGACCGGCCCCCCCC 96549

QY	1	GGACCCACCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTTACCCACCCACCTC	60
DB	5260	GGACCCACCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTTACCCACCCACCTC	5319
QY	61	ACACCCACCAACCCACCAACCTCTCTCTCTCACCGCGGACCGGCGCCACCCACCCACCC	120
DB	5320	ACACCCACCAACCCACCAACCTCTCTCTCACCGCGGACCGGCGCCACCCACCCACCC	5379
QY	121	CACGACCAACACTCACCCACCACTCCAAACAAAAGGATCCACCTCACCATCACCCACC	180
DB	5380	CACGACCAACACTCACCCACCACTCCAAACAAAAGGATCCACCTCACCATCACCCACC	5439
QY	181	TCGACACACAGCA-ACCCAGACCACTC-CAACAATC-CTCAAACGATCCCGCCACAA	237
DB	5440	TCGACACACAGCAACCCAGACCACTCAACAACCTCACTCAACACCACTCCCGCCACAA	5499
QY	238	CACCCCTCATCACCGCTCATCACACCGGAGGATCTCTGACGAGCGACCCCTTACCCAC	297
DB	5500	CACCCCTCATCACCGCTCATCACACCGGAGGATCTCTGACGAGCGACCCCTTACCCAC	5559
QY	298	CTCACCCCCACCAACTCAACACCGTCTCTCGCGCCAAAGCCACAGCGCCCACTCTCTC	357
DB	5560	CTCACCCCCACCAACTCAACACCGTCTCTCGCGCCAAAGCCACAGCGCCCACTCTCTC	5619
QY	358	CACCAACTCACCAACACCCCTCTCACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTC	417
DB	5620	CACCAACTCACCAACACCCCTCTCACCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTC	5679
QY	418	ACCTTCGGCGGACCGGCGCAACCACTACGCGCGGACCAACGCGCTTCTCTCTCTCTCTC	477
DB	5680	ACCTTCGGCGGACCGGCGCAACCACTACGCGCGGACCAACGCGCTTCTCTCTCTCTCTC	5739
QY	478	GCCCAACCAACCCACCAACCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	534

Db	5740	GGCCACCACCGCACACCCACCTCCCGCCACCGACGATCGCTCGGGGACCTGG	5796
RESULT 5			
AB070940/c			
LOCUS	AB070940	104326 bp	DNA linear BCT 22-JAN-2002
DEFINITION	Streptomyces avermitilis oligomycin biosynthetic gene cluster.		
ACCESSION	AB070940		
VERSION	AB070940.1	GI:15823967	
KEYWORDS	Streptomyces avermitilis DNA.		
SOURCE	Streptomyces avermitilis		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)		
AUTHORS	21477403		
TITLE	2 (bases 1 to 104326)		
JOURNAL	Direct Submission		
MEDLINE	Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)		
AUTHORS	Ikeda, H.		
JOURNAL	Location/Qualifiers		
FEATURES	1..104326		
source	/organism="Streptomyces avermitilis"		
gene	/db_xref="taxon:33903"		
CDS	complement(1..1134)		
	/gene="mcma"		
	complement(1..1134)		
	/gene="mcma"		
	/note="methylmalonyl-CoA mutase"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="BAB69185.1"		
	/db_xref="GI:15823968"		
	/translation="MTGISTADANARFQLVSGSTRAVVPVDFLPTRAGLSDAPAA RGLVRRGVAVDSIDDMRVLFGLRDLRVFSLVADPAAPLLVLQVLSVBHGAQAQ RLAVRGDVFKEFLARGFVFPPLRPSLRAMNFAYCQAEPRWCQSVCCGHLAGT GADPALEAVFTVAGGIEYLRAAVTAVPRLTCLVGGTTGRENRAKLRTTR VVTRTANERFGSGRTAPRLYVTRPRDGLATLACTEDIEAGATLVDRVIRLIGTTLVL EHLAQLGRPHGDRPATPRPHPHQDSLDLQASQTERLAKLRWRVQPAVDEALLR LRKAQGDNDVLYPNRELAARATVGEVCDALURSTWGTGSPSAT"		
gene	complement(1131..1886)		
CDS	/gene="ppta"		
	complement(1131..1886)		
	/gene="ppta"		
	/codon_start=1		
	/transl_table=11		
	/product="phosphopantetheinyl transferase"		
	/protein_id="BAB69186.1"		
	/db_xref="GI:15823969"		
	/translation="MRAHEVFADVPVSLHPREALVODAVESRRERFATGRHCARRA LASLGVPPGLLATPRGPRWPGVAGTICGACGYRAAARAPGIRAVGIDAEIDRPL LPEGVLEVALPAEQFQVKEHTARHPGVWVGKLLFSAKESVYKAWFPATRGPLGFEA LIDLPVSSSTFRVRLPTALRAABPELDLHGRMLVRHGIIVTAVVVDGPGSDTTPHAP SPYDRTKHSAPTGRSADPGRGPTDRATPTA"		
gene	complement(1910..2875)		
CDS	/gene="simX4 homolog"		
	complement(1910..2875)		
	/gene="simX4 homolog"		
	/codon_start=1		
	/transl_table=11		
	/product="SimX4 homolog"		

THE KITASATO INSTITUTE
OS Streptomyces avermitilis
PN JP 2000245457-A/1
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR SATOSHI OMURA, HARUO IKEDA
PI C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
C1201/68/CO70493/22,
PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (11..(11916)
FT CDS (11971)..(30687).
FEATURES
source
1..30690
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 5355 a 12455 c 8617 g 4263 t
ORIGIN
Query Match 100.0%; Score 534; DB 6; Length 30690;
Best Local Similarity 100.0%; Pred. No. 4.1e-58;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGAACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 60
DB 5143 GGAACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5202
QY 61 ACCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 120
DB 5203 ACCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5262
QY 121 CAGGCAACACCTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 180
DB 5263 CAGGCAACACCTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5322
QY 181 TCGGACACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 240
DB 5323 TCGGACACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5382
QY 241 CCGCTCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 300
DB 5383 CCGCTCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5442
QY 301 ACCCCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 360
DB 5443 ACCCCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5502
QY 361 CAACTCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 420
DB 5503 CAACTCACCACTCATCACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5562
QY 421 TTCCGCGGACCGCGGCAACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 480
DB 5563 TTCCGCGGACCGCGGCAACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5622
QY 481 CACCACCGGACCGCGGCAACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 534
DB 5623 CACCACCGGACCGCGGCAACCGCGGAACCGCGGCTCGCCACCCACCTCCACCCACCACTCC 5676
RESULT 2
AB032367
LOCUS
DEFINITION Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION AB032367
VERSION
KEYWORDS AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVE5 4; type I polyketide synthase AVE5 3; type I polyketide synthase AVE5 2; type I polyketide synthase AVE5 1.
SOURCE Streptomyces avermitilis
ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 64957)
Ikeda, H., Nonomiyama, T., Usami, M., Ohta, T. and Omura, S.
Organization of the biosynthetic gene cluster for the polyketide anthelmintic macroide avermectin in Streptomyces avermitilis
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
99380548
2 (bases 1 to 64957)
Ikeda, H., Nonomiyama, T., Usami, M., Ohta, T. and Omura, S.
Direct Submission
Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial Chemistry; S-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)
Location/Qualifiers
1..64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
101..12019
/gene="aveA1"
101..12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2"
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVE5 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDGGEPRPAAGEVLGVADADGGVVFVPPQGGPQWQPMGR
ELLASDVRESVRACEAPAPYVDMSVEQVLRDSDPAGLDVDDVQPTLPVAMISL
AALWRSQVEFCVGLHSGEIAAHSVSGSLADAAVTVLWSAQTLTLAGTGLVLS
VAATPELLPRTIAPWTEINPARLAAVAVNGPRSTVVGSAVADLVADLTAQVTR
MIPVDPAHSPLMYAIEERVSGLLPITPRSPRIPFSSVSGRDLDELDAAYWRN
MSGTVPFPAALLLQQGTFFVMSHPVLTGMQLQELAPDLGDTTGTADVTIMGLR
RGQGLDHFLLSLAQHGHGETSATVLSARLTALSPQQOQLLDLVAHRAHMAVLD
DGNERTADGASASPAHLGDSVNGVELNRSLKATGLPVTLLI FPHHTPAVAAAR
LRTAALHLDDEDTAPVPSPSGCGTAAADDPDIIAGNACFPFGVSPKOLWELAA
GGDAIAGPPTDORHQAQDPTGPTFGQGGFLHDAHFDAGFFGSLPRALLA
MDPQRLLETSWEAFRAGIDPLSVGRSGTGFAGALSFDYGRMDTSFSGEADVE
GHILTTGSLGRISAYSGLEGPATVDTGCSASVTLHLACSLRSBECTALAG
GVSMSTLGMFIEFSRQSKAYSAADGTGEGVGLLVERLSDAVRLGH
RVLAIVRGSVAVNGASGLTAPNGPAQERVIRQALANAGLSVADVVEGHTCTTL
GDPIEAQALLATYGRAGDRFLWGLSKSNTGHTMAAGGVGVIKNWALREGVLPRT
LHVDPSPQVMSAGAVLLTEAVPMDGAGLRRLRAGVSFGIGGTNAHVILSEAPA
AGCGVGGVLEGAPGLAISVAESVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV
LRAQAEALQYVAVRVDVSLADVGAGLACRAVLEHRAVLAADREELVQGLGALAAG
EPDRVTTGAPGGVGVVFPVGGGOWAGMGLVLLASGVPFARRMOACBEALAPW
VDMVVDILREDDAGDAVVERADVQVPLPSVMSIALALWRSYIEPDAVLHSGEIA
AAHVCGALSXDAKATVALSRALAANVRGGMASVPLPAQVEQLIGERNAGRLWA
AVNGRSTAVSGDAEAVDEIAYCAGTGVARRIPVDIYASHCPVQVPRELLELGD
ISPOSGVFFETVEGTWDTTLLDAAYVYRNLHLPVPSDAVQALADGHRVFEVS
PHPTLVPAIEDTETDAEDVTIAGSLRGDNDTRFLTALAHHTTGTGTPTTHHHY
THHTPHPHTHLDLPYFQHOHVLESGQSGAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS
EVESEFMDAARQLETVATVAVPPPSAGLTVVPAVSAHHRHODQARINTWYQET
WKPLTPTTHQPHQTLTATPETQTHPHITNLTLNHHGHTPIPLTLNHTHTNPHQ
LHHTLHTRQANHTTGAITGLSLDALDELPHPHHTPTGLTLNLTQTQTHQTH
PPTPLWYATNTATNTDPLTHPTQAGTGLARLTLEHTHTAGIIDLPTPTPT
LOHLTQTHQPHQTLTATPETQTHPHITNLTLNHHGHTPIPLTLNHTHTNPHQ
ATHLTHTHTQTHQTLTATPETQTHPHITNLTLNHHGHTPIPLTLNHTHTNPHQ
QLLNTIPQHPTLTVHTAGILDDATLNLFTPLNLYRAKASHALLHOLTOHTPL
TAFVLSAAATFAGPQNTAAANAYDALAHRRHTHLPATSIANGTQNGSLADS
DKARAYLDRRGRFRMSPELATAVQIADTERPVYVIADIDWSKIEHTQSDGLVS
AREREPAVQRTPEPAELHKLHATQSDARALHLEHRAVLAADREELVQGLGALAAG
FRALGDSLTAVEFENLLIKATGLRPSVLFDPHETPAKLAHLQNRQARAAEAPS
AAVTAESVETPIAVGMACRFQGVTSADDVMDLISEQDAIGGPTDRGWMDTL
YDPDHDGTCYTRNGGFLYDAGHDFAEFGLISPREALMDPOORLLLETWETLEHA
GPNHTLHGTCTGT
GPVSVDTACSSSLVALHACALRAGECMALAGGVTVMSPSGAVSFBSRQGLAAD
GCKAFSAADAGTGWGEGVGLLVERLSDAHNRHRLVAVRGSVAVNGDAGLTAAP

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1035.85 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676

Perfect score: 534

Sequence: 1 ggaacacacctcatcaccgg.....gcatgcctggggcaccctgg 534

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sta.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sta.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_higo_hum.*
- 40: em_higo_mus.*
- 41: em_higo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	534	100.0	30690	6	E38020	E38020 Avermectin
2	534	100.0	64957	1	AB032367	AB032367 Streptomycin
3	532.4	99.7	12381	6	AX006889	AX006889 Sequence
4	497.8	93.2	11096	1	AF275943	AF275943 Streptomycin
5	335.6	62.8	104326	1	AB070940	AB070940 Streptomycin
6	285.8	53.5	41097	1	AF016585	AF016585 Streptomycin
7	266.8	50.0	24225	1	SC2C4	AL512902 Streptomycin
8	266.8	50.0	123580	1	AF263912	AF263912 Streptomycin
9	266.8	50.0	125401	6	AX211739	AX211739 Sequence
10	260.4	48.8	65140	6	AX211705	AX211705 Sequence
11	257.4	48.2	19791	1	SPFKBAD	Y10438 Streptomycin
12	252.6	47.3	28732	1	AF082100	AF082100 Streptomycin
13	246.8	46.2	32870	1	AF007101	AF007101 Streptomycin
14	244.4	45.8	113193	1	AF357202	AF357202 Streptomycin
15	244	45.7	77534	1	AF235504	AF235504 Streptomycin
16	242.4	45.4	77534	1	AF235504	AF235504 Streptomycin
17	236.4	44.3	30000	6	AX250262	AX250262 Sequence
18	234.8	44.0	84985	1	SNA278573	AJ278573 Streptomycin
19	226.8	42.5	20394	1	SNA132222	AJ132222 Streptomycin
20	226.8	42.5	20394	6	AX067996	AX067996 Sequence
21	225.2	42.2	27541	6	AX211706	AX211706 Sequence
22	225.2	42.2	104326	1	AB070940	AB070940 Streptomycin
23	223.6	41.9	37948	1	AF079138	AF079138 Streptomycin
24	217.4	40.7	31422	6	E38021	E38021 Avermectin
25	217.4	40.7	64957	1	AB032367	AB032367 Streptomycin
26	214	40.1	3573	1	AB016763	AB016763 Streptomycin
27	212.4	39.8	26195	1	SC1G7	AL591083 Streptomycin
28	210.8	39.5	30000	6	AX250263	AX250263 Sequence
29	210.8	39.5	43280	1	SFU78289	U78289 Streptomycin
30	207.6	38.9	49377	6	I88042	I88042 Sequence 1
31	207.6	38.9	67523	1	SCU24241	U24241 Sorangium c
32	201.2	37.7	47981	1	AF263245	AF263245 Micromono
33	201.2	37.7	47981	6	AX112026	AX112026 Sequence
34	200.8	37.6	107379	1	SHGCP1R	X86780 S.hygrosco
35	199.6	37.4	39314	1	SGR300302	AJ300302 Streptomycin
36	196.4	36.8	28958	6	AR044578	AR044578 Sequence
37	196.4	36.8	28958	6	I47768	I47768 Sequence 6
38	196.4	36.8	28958	6	I50958	I50958 Sequence 6
39	196.4	36.8	28958	6	I63356	I63356 Sequence 6
40	196.4	36.8	28958	6	I70387	I70387 Sequence 6
41	196.4	36.8	28958	6	I85639	I85639 Sequence 6
42	196.4	36.8	28958	6	I88045	I88045 Sequence 4
43	196.4	36.8	28958	6	I90322	I90322 Sequence 6
44	196.4	36.8	41782	1	AX118081	AX118081 Streptomycin
45	195.2	36.6	82746	1	AF453501	AF453501 Actinosyn

ALIGNMENTS

RESULT 1

E38020

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

E38020 30690 bp . DNA linear PAT 31-JAN-2002
Avermectin aglycon synthase gene.
E38020.1 GI:18626909
JP 2000245457-A/1.
Streptomyces avermitilis.
Streptomyces avermitilis.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
1 (bases 1 to 30690)
Omura,S. and Ikeda,H.
Avermectin aglycon synthase gene
Patent: JP 2000245457-A 1 12-SEP-2000;

THIS PAGE BLANK (USPTO)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 4832 row: 0 column: 2

Seq primer: T7

Class: BAC ends

High quality sequence stop: 918.

FEATURES

source

1. .918

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate:4832 Col=2 Row=0"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 16 a 93 c 718 g 60 t 31 others

Query Match

Best Local Similarity 8.0%; Score 73.6; DB 17; Length 918;

Matches 287; Conservative 0; Mismatches 345; Indels 3; Gaps 1;

QY 2 TCTTCGTGTTCCGACACAGGTGGCAGTGGCCCGGGATGGGTGTCGTCTGCTGCCT 61

DB 262 TCTTTGGGNTTGG 321

QY 62 CCTCTCCGTTTCGCCCGCGGATCGAGCGTGCAGAGAGGCTCTGGCGCCGTGGGTGG 121

DB 322 GCGGNGGG 381

QY 122 ACTGTGTCGTGTGGACATCTCTGCGCCGGACCGCGGGGATGCGGTGTCGGAGCGGGCCG 181

DB 382 GCGGGCGGG 441

QY 182 ATGTGTCCAGCTGTGCTGTTCAGCGTCATGTTGCTCTTGTGCTGCTCTGTGCGTTCT 241

DB 442 GGGGGG---GCCCGGG 498

QY 242 ACCTATCGAACCCGACGGTCTTGGGCGATTCCAGGGCGAGATCGCGCGCGCGCATG 301

DB 499 GGG 558

QY 302 TGTGTGGGCGCTGAGCGTGAAGACCGCGCGAAGACTCTTGGCTGCGCAGCGCGGCG 361

DB 559 GNGGG 618

QY 362 TGGCCGCTGTGGGGGGCCCGGGCGGCATGCGCTCAAGTGCCTGCTGCCAGAGGTGG 421

DB 619 GCGGG 678

QY 422 AGCAGCTCATTTGAGTGGTGGCGGGCGGGCGGTGTTGGGTGGCGCGGTCAACCGCCCC 481

DB 679 GGG 738

QY 482 GCTCCACACCGCTCTCGGGGGATCGGAGCGGTGTCAGAGTGTGCGTACTGTGCG 541

DB 739 GGG 798

QY 542 GCACCGGGGTGCGGGCCCGCGGATCCCGGTGCGACTATGCTTCGCACTGCCCCCATGTGC 601

DB 799 GCGCGGG 858

QY 602 AGCCCTCTCGGGAGGAGTTGCTGGAGCTGCTGGGG 636

DB 859 GGGCGGG 893

Search completed: June 18, 2003, 01:00:14

Job time : 1183.72 secs

Best Local Similarity 47.9%; Pred. No. 0.00055;
Matches 268; Conservative 0; Mismatches 289; Indels 3; Gaps 2;

QY 14 CCGACAGCGTGGGAGTGGCGCGGATGGGTGCTGCTCGCTCTCTCCGGTGT 73
Db |||||
374 CCGGGGGGGGGCGCGGCGGNGCGCGCGCGGGGGGGCGGTTTGTGCGCCNGGGG 433
QY 74 TCGCCCGCGCGATGACAGGCGTGGAGAGGCTCTGCGCGCTGGGTGGACTGTCTGTGG 133
Db |||||
434 GGGGGGGCGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 493
QY 134 TGGACATCTGCGCCCGGACGCGGGGGATGCGGTGTGGAGCGGGCGGATGTGTCCAGC 193
Db |||||
494 GCGAGAGGG 553
QY 194 CTGTGCTGTTCAGCGTCGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAAC 253
Db |||||
554 CCGGGGGGGGGCGGG 613
QY 254 CCGACCGCGTCTTGGCCATTCACGAGCGAGATCGCGCGCGCATGTGTGGGGC-- 311
Db |||||
614 GGG 673
QY 312 GCTGAGCCTGAAGACCGCGCGAAGACTGTTCGCTGCGCAGCCGGCGCTGCGCGCTGT 371
Db |||||
674 GCGGGGGGGGGCGGACCGGTCCGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 733
QY 372 GCGGGGGGGGGCGGATGCGCTCAGTGGCGTCTGCGCAGCGCGCGCTGCGCGCGTGT 431
Db |||||
734 CCGGGGGGGGGCGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 793
QY 432 TGTGAGCGTGGCGGG 491
Db |||||
794 GGGGGGGGGTGCAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 852
QY 492 CGTCTCGGGGGATGCCAGCGCGTGGACGAGTGTGCGGTACTGTGCGGACCGCGGGGT 551
Db |||||
853 GGGGGGGGGGGAGGG 912
QY 552 GCGGGGGGGGGCGGATCCCGG 571
Db |||||
913 GGGGGCGCGCGCGCGGGGG 932

RESULT 14
BG441241/c
LOCUS
DEFINITION
GA_Ea012G10f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea012G10f, mRNA sequence.
ACCESSION
BG441241
VERSION
BG441241.1 GI:13350893
KEYWORDS
EST.
SOURCE
Gossypium arboreum.
ORGANISM
Gossypium arboreum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE
1 (bases 1 to 925)
Wing,R.A., Friebe,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence start: 39
High quality sequence stop: 903.

FEATURES
source
Location/Qualifiers
1..925
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea012G10f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 57 a 687 c 133 g 42 t 6 others
ORIGIN
Query Match 8.0%; Score 73.8; DB 12; Length 925;
Best Local Similarity 45.9%; Pred. No. 0.0006;
Matches 252; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

QY 16 GGACAGCGTGGGAGTGGCGCGGATGGGTGCTGCTGCTCGCTCTCTCCGGTGTTC 75
Db |||||
773 GGGGGTGG 714
QY 76 GCCCGCGGATGCAGCGTGCAGAGAGCTCTGGCGCGCTGGGTGGAATGTGTCTGTGTG 135
Db |||||
713 GGGCGTGG 654
QY 136 GACATCTGCGCGCGGACCGGGGGATGCGGTGTGGAGAGCGGGCGGATGTGTCCAGCT 195
Db |||||
653 GGGGGCGGG 594
QY 196 GTGCTGTTACGCTCATGCTGCTTTTGGCTGCTCTGTGCGCTTCCTACGATCGAACCC 255
Db |||||
593 GGGGGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 534
QY 256 GACGCGGCTCTTGGCGATTCCACGCGGAGATCGCGCGCGGATGTGTGTGGGGCGGTG 315
Db |||||
533 GGG 474
QY 316 AGCTGTGAAGACCGCGCGAAGACTGTTGGCTGTCGACAGCGGGCGCTGGCGCTGTGCGG 375
Db |||||
473 GGG 414
QY 376 GCGCGGGCGCGCATGCGCTCAGTGCCTGCTGCCAGGAGTGGAGCAGCTCATTTGT 435
Db |||||
413 AGGCGGGCGCGGG 354
QY 436 GAGCGGTGGCGGGCGGTTGTGGTGGCGCGGTCAACGCGCCCGCTCCACCGCGGTTC 495
Db |||||
353 GGGGGTGGCGGG 294
QY 496 TCGGGGGATCGGAGCGGTGACAGGTGCTGGGCTACTGTGCGGACCGGGGTGCGG 555
Db |||||
293 GGGAGGG 234
QY 556 GCCCGCGCG 564
Db |||||
233 AGGGGGGGGG 225

RESULT 15
AQ895329
LOCUS
DEFINITION
HS_4832_A2_H01_T7A CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=4832 Col=2 Row=O, DNA sequence.
ACCESSION
AQ895329
VERSION
AQ895329.1 GI:6351435
KEYWORDS
GSS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 918)

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES Location/Qualifiers
source 1..1101

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN12114"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end: 77"

BASE COUNT 83 a 566 c 179 g 93 t 180 others

Query Match 8.1%; Score .75; DB 17; Length 1101;

Best Local Similarity 42.2%; Pred. No. 0.00036; Indels 0; Gaps 0;

Matches 237; Conservative 25; Mismatches 299; Indels 0; Gaps 0;

QY 76 GCCCGCGGATGCGGTCGAGAGGCTCTCGCGCGGTGGTGGACTGGTCTGTGGTG 135

Db 608 GCGGGGGGGGNSGCGNSCGCGCGGGGCGCCSCCGCGCGGGGGGGG 549

QY 136 GACATCTGCGCGGACGCGGGGATGCGGTGTGGAGCGGCGCATGTGTTCAGCTT 195

Db 548 GCGGGGGGGGCGGGGCGCGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGG 489

QY 196 GTGCTGTTACGCTCATGTGTTGTTGGTCTCTGTGGCTTCTACGATATCGAACCC 255

Db 488 CCGGG 429

QY 256 GACCGGCTCTTGGCCATTCCAGAGCGAGATCGCGCGCGCATGTGTGTGGGGCGCTG 315

Db 428 GCGCGGGCGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 369

QY 316 AGCCTGAAGACCGCGGGAAGATGTGTGCTCTCGCAGCGCGCGCTGCGCGTGTGCGG 375

Db 368 GCGGG 309

QY 376 GCGCGGGGGGCGATGCTTCTCAGTCCGCTGCTGCTGCGCGCGAGGTGGAGCTATTGCT 435

Db 308 GGG 249

QY 436 GAGCGGTGGCGGG 495

Db 248 GGG 189

QY 496 TCGGGGGATCCGAGCGGTGAGAGAGTCTGCGGTACTGTGCGGACCGGGGTGCGG 555

Db 188 GCGGG 129

QY 556 GCCCGCGGATCCCGGTTCGACTATGCTCGCACTGCCCCCATGTGAGCCCCCTGCGGGAG 615

Db 128 GGG 69

QY 616 GAGTTGCTGAGCTGTGCGG 636

Db 68 GGGCGGGCGGSGGGGGGGGG 48

RESULT 11

BQ670959

LOCUS

DEFINITION BQ670959 997 bp mRNA linear EST 15-JUL-2002

AGENCOURT 8203742 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255995

5', mRNA sequence.

ACCESSION BQ670959

VERSION BQ670959.1 GI:21781793

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 997)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM2407 row: p column: 12

High quality sequence start: 156

High quality sequence stop: 281.

FEATURES

source

Location/Qualifiers

1..997

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6255995"

/clone_lib="NIH_MGC_102"

/tissue_type="epidermoid carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: salivary gland; Vector: pOTB7; Site: 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Library constructed

by Ling Hong in the laboratory of Gerald M. Rubin

(University of California, Berkeley) using ZAP-CDNA

Synthesis Kit (Stratagene) and Superscript II RT (Life

Technologies). Note: this is a NIH MGC Library."

BASE COUNT 56 a 211 c 579 g 77 t 74 others

ORIGIN

Query Match 8.1%; Score 74.8; DB 14; Length 997;

Best Local Similarity 45.7%; Pred. No. 0.00039;

Matches 245; Conservative 0; Mismatches 288; Indels 3; Gaps 1;

QY 21 GGGTGGCGAGTGGGCGGATGGGTGTGCTGTCTCGCTCTCTCGCTGTTCGCCCG 80

Db 465 GGG 524

QY 81 GCGGATGCGCGTCCGAGGAGGTCTGCGCGCGGTGGTGGACTGGTCTGTGGACAT 140

Db 525 GGG 584

QY 141 CCTGCGCGGAGCGGGGGATGCGGTGTGGAGCGGCGCGATGTGTGCTTGTGCT 200

Db 585 NGCGGG 644

QY 201 GTTACGCTCATGTGTCTTTTGGTGTCTTGTGGGTCTTCTACGCTATCGAACCG 260

Db 645 GGG 704

QY 261 GGTCTCTGGCCATTCCAGGCGAGATCGCGCGCGCGCATGTGTGTGGGGCGCT 320

Db 705 GG---GG 761

QY 321 GAAGACGCGCGGAAGACTGTGCGCTGCGCGAGCGCGGGCGCTGCGCGCGCGCG 380

Db 762 GGG 821

QY 381 GGGCGCATGGCTCAGTGTGCGCTCCCTGCGCGAGGTGGAGAGCTCATTTGGTG 440

Db 822 GGG 881

QY 441 GTGGCGGG 500

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM1951 row: f column: 06

High quality sequence stop: 379.

FEATURES

.source

1. .1293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5455109"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 71 a 509 c 495 g 54 t 164 others

ORIGIN

Query Match 8.5%; Score 78; DB 14; Length 1293;

Best Local Similarity 40.4%; Pred. No. 0.0001;

Matches 333; Conservative 0; Mismatches 482; Indels 10; Gaps 2;

QY 91 GCGTGGAGAGGCTCTGGCGCGTGGTGGACTGTCTGTGTGACATCTCTGCGCGG 150

DB 1244 GCGGGGNGCGGGGGCGGGGGCGCGGCGCGCGGGCGGGGGCGGGGGCGGGGG 1185

QY 151 GAGCGGGGATCGGTGTGGAGCGGGCGGATGTGTTCAGCGTCTGTCTTCAGCGTC 210

DB 1184 GCGCGCGGNGGCG 1125

QY 211 ATGGTGTCTTTGCTGTCTGTGGGTCTTCTACGATATGAAACCGACGCGTCTTGGC 270

DB 1124 GNGCGCGGNGCGGGGGCGGGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 1065

QY 271 CATTTCCAGGCGGATCGCGCGCGCGCATGTGTGTGGGGCGTGTGAGCTGAAGACGG 330

DB 1064 CGCGCGGNGNGGGGGCGGGGGCGGGGGCGGCGCGCGCGCGCGCGCGCGCGCGCG 1005

QY 331 GCGAAGACTTTGCTGTGCGTGCAGCGCGCGCTGTGTGCGGGCGGGCGGCATG 390

DB 1004 GGG-----GGGGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 954

QY 391 GCCTCAGTCCGTCTGCTGCCAGAGGTGAGAGCTCATTTGTTAGCGTGTGGCGGG 450

DB 953 GGGNCGCGGGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894

QY 451 CGGTTGTGGTGGGGGGTCAACGGCCCCCGCTCCACCGCGCTCTCGGGGATCGGAG 510

DB 893 GCGGGG-GGGCGGGGNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 835

QY 511 GCGGTGGACAGGTGTGGCGTACTGTGCGCGCACCGGGGTGCGGGCGCGCGCGATCCCG 570

DB 834 GGGGCGCGCGGGGNGCGCGCGCGGGCGCGCGCGGGGGCGCGGGGGGGGGGGGG 775

QY 571 GTCAGTATGCTTCGCACTGCCCCCATGTGACGCCCCCTGCGGAGAGTGTCTGAGCTG 630

DB 774 CGCGCGCGCGGGGGCGGGCGGGCGGCGGCGGCGGGGGGGGGGGGGGGGGGGGG 715

QY 631 CTGGGGGACATACCGCGAGCGTCCGGCGTCCGCTTCTTCTCCAGTGTGAGGGCACC 690

DB 714 GNGGGGCGGNGGGGGGGGGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 655

QY 691 TGGCTGGACACCAACCCCTGGACGCGCGCTACTGTGTACCGCAACCTGACACGCGGTC 750

DB 654 CCGGNGGGCCCCCGCGCGGNGGCGCGCCCGGCGCGCGCGCGCGCGCGCGCGCG 595

QY 751 CGTTTACGCGATCGCGTTCAGAGCGCTTGGCGGATGACGGACACCGCTTCTTCGAGTC 810

DB 594 CCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 535

QY 811 AGCCCCACCCACCCCTCGTCCCGCATCGAAGACACACCGAAGACACCGCGGAGAC 870

DB 534 CCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 475

QY 871 GTCACCGGATCGGCGAGCTCTCGCGCGCGGCGGACAAACGACACCGCG 915

DB 474 NNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 430

RESULT 8

AG126210 842 bp DNA linear GSS 04-NOV-2001

LOCUS Pan troglodytes DNA, clone: PTB-136L05.R, genomic survey sequence.

DEFINITION AG126210

ACCESSION AG126210.1 GI:1665375

VERSION GSS.

KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

SOURCE BAC Library clone: PTB-136L05.R.

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

1

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 842)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chimbes@sc.riken.go.jp, URL: http://ngp.gen.riken.go.jp/;

Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

clones are generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .842

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-136L05.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT 32 a 214 c 479 g 14 t 103 others

ORIGIN

Query Match 8.2%; Score 75.4; DB 17; Length 842;

Best Local Similarity 43.0%; Pred. No. 0.00031;

Matches 253; Conservative 0; Mismatches 336; Indels 0; Gaps 0;

QY 16 GCACAGGTGGGACGTGGGCGCGGATGGGTGTGCGTCTCTCTCTCTCTCTCTCTCT 75

DB 233 GGGGNGGG 292

QY 76 GCCCGCGGATGCAAGCGGTGCGAGAGAGCTCTTGGCGCCGCTGGGTGGACTGTGTGTG 135

DB 293 CCNGGG 352

QY 136 GACATCTCTCGCGGAGCGGGGGGATCGGTGTGGAGCGGGCGGATGTGTTCAGCCT 195

DB 353 GGG 412


```

Db      433 GGCCTGGGGGGGGGGGNGCGCGCCGGGGGGGGNNCGGGGNGGCGCGGGGGGGGGCGCGC 492
QY      449 GCGCGTTGTGGTGGCGGGCGGTCAACGGCGCCCCCGCTCCACCGCCGTCCTCGGGGGATGCCG 508
Db      493 GCGCGGGGGGAGGGGGGGGGCGGGCGGGCGGGCGGGCGGGCGGGGGGGGGGGGGGGGGGG 552
QY      509 AGCGGTGACGAGGTGCTGGCGCTACTGTGTCCCGCACCGGGGTCGGGGCCCGCGCATCC 568
Db      553 GGGGGCCCCCGCGGGGGCGGGGNGCGGGGCGGGCGGGCGGGCGGGGGGGGGGGGGGGGG 612
QY      569 CGGTCACTATGCTCGACTCGCCCTCCCATGTGAGCCCTTGGCGGAGAGTGTCTGGAGC 628
Db      613 GCGGGGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 672
QY      629 TGTCTGGGGGACATCAGCCCGCAGCGCTCGCGGC 660
Db      673 GCGCGGNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 704

RESULT 3
LOCUS   CNS00720          932 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL066742
VERSION   AL066742.1 GI:4945205
KEYWORDS
SOURCE    GSS.
ORGANISM  Drosophila melanogaster.
            Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 932)
REFERENCE
AUTHORS   Aaron Mammoler in Pieter de Jong's laboratory in the Department of
TITLE      Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL    NY. The library is named RPCI-98 and was constructed by partial
COMMENT    EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            p1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
            Location/Qualifiers
FEATURES             1..932
                    /organism="Drosophila melanogaster"
                    /db_xref="taxon:7227"
                    /clone="BACR14B09"
                    /clone_lib="RPCI-98"
                    /note="end : T7"

BASE COUNT  155 a  202 c  241 g   91 t   243 others
ORIGIN

Query Match      8.9%; Score 82.2; DB 17; Length 932;
Best Local Similarity 32.5%; Pred. No. 1.8e-05;
Matches 161; Conservative 113; Mismatches 219; Indels 2; Gaps 1;

QY      174 GCGGCGCGATGGTCCAGCCGTGTGTTCAGCGTCATGGTCTTTGGCGTCTGTG 233
Db      419 GCGSTCKCKGCKCTGCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 478
QY      234 GCGTTCTTACGGTATCGAACCAGCGCG--GTCTTGGCCATTCCAGGCGGAGATCGCG 291

```

```

Db      479 GTKBCGKGKCGKCKGCGCGGGGGGVCGRGGSGSGSCGTTGGCGGKGCGCKGCTCS 538
QY      292 GCGCGCGCATGTGTGTGGGGCGCTGAGCCCTGAAGGACGGCGGCAAGACTGTTGCGCTGCGC 351
Db      539 KGGCTGBGSTGCGGGGGSKSGSGGSCBGC CGGGGSTKGGSGGGGGBSBSTKTBKKGK 598
QY      352 AGCCCGGGCGCTGCGCGCTGTGCGGGCGCGGGCGGCATGCGCTCACTGCGCTGCGCTGCC 411
Db      599 CGSSSGSCSGSGSCHRGGGGCGSGCGSGCGGGSSCGSCSCGCGSGGGCGCGCGSG 658
QY      412 CAGGAGGTGAGACACTCATTTGTGAGCGGTGGGCGGGCGGCTTTGTGGTGGCGGCGTTC 471
Db      659 SSGSGSGSGSGSGSGCGCGSCCGCGSCCGCCSCCGSCCGCGCGCGCGCGCGCGCG 718
QY      472 AACGGCCCCCGCTCCACCGCGCTCTCGGGGGATGCGGAGGGGTGGAGCGAGTGTGGCG 531
Db      719 SCGSGSGSGCGSCCGSCCGSCCGSCCGSCCGSCCGSCCGSCCGSCCGSCCGSGSG 778
QY      532 TACTGTGCGCGCACCGGGTGTGCGCGCGCGCGGATCCCGGTGCGATATGCTCGCACTGC 591
Db      779 CGSCGCGCGSGSGSGCGSGSGSGCGSGSGSGCGSGSGSGCGSGSGSGCGSGSGSGSG 838
QY      592 CCCCATGTGACAGCCCTCGCGGAGAGTTGTGTGAGTCTGTGGGAGACATCAGCCGCGAG 651
Db      839 CCCCSSGCGCGCGSGSGCGSGSGCGSGSGCGSGSGCGSGSGCGSGSGSGSGSGSGSG 898
QY      652 CGGTCTCGCGGTGCGCG 666
Db      899 GCSSGCGSGCGCGSS 913

RESULT 4
LOCUS   BM810045          1364 bp      mRNA      linear      EST 05-MAR-2002
DEFINITION AGENCOURT 6579545 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5469986
            5', mRNA sequence.
ACCESSION BM810045
VERSION   BM810045.1 GI:19126868
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1364)
REFERENCE   NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILLNL at:
            http://image.llnl.gov
            Plate: LLCM1976 row: b column: 03
            High quality sequence stop: 178.
            Location/Qualifiers
FEATURES             1..1364
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5469986"
                    /clone_lib="NIH_MGC_41"
                    /tissue_type="amelanotic melanoma, cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                    EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                    into EcoRI/XhoI sites using the following 5' adaptor:
                    GGCACGAG(G). Library constructed by Ling Hong in the
                    laboratory of Gerald M. Rubin (University of California,
                    Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

```


/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:546986"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GSCACAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT 40 a 356 c 673 g 27 t 268 others
ORIGIN

Query Match 9.8%; Score 90.4; DB 14; Length 1364;
Best Local Similarity 44.9%; Pred. No. 5.9e-07;
Matches 301; Conservative 0; Mismatches 362; Indels 7; Gaps 2;
QY 249 CGAACCCGACGGGCTTGGCCATTCCAGGCGGAGATCGGGCGGCATGTGTGG 308
DB 1255 CGNCCGCGCGCCCGCCGCGGGGCGCCGCGCGCGCGCGCGCGCGCGCG 1196
QY 309 GCGCGCTGAGCTGAAGGACGCGCGCAAGACTGTGTGGCTGCGCAGCGCGCGCTGGCGCG 368
DB 1195 GCGG-GGCG 1137
QY 369 TGTGCGGCGCGCGCGCGCATGCGCTTCAGTGGCGCTGCTGCGCGCGCGCGCGCGCGCGCG 428
DB 1136 NGCG 1077
QY 429 CATTGTGTGACGTGGCGCGCGCGGTGTGGTGGCGCGGTCAAGCGCGCGCGCTCCAC 488
DB 1076 CGGCG 1017
QY 489 CGCGCTCTCGGGGATGCGCGAGCGGTGGACGAGGTGCTGGGCTACTGTGCGCGCGCACCG 548
DB 1016 GCG 957
QY 549 GTGCGGCGCGCGCGATCCCGTTCGACTATGCTTCGCACTGCGCCCTATGTGACGCGCT 608
DB 956 CG 897
QY 609 GCGGAGGAGTGTCTGGAGTGTGGGAGCATCAGCCGCGCGCGCGCGCGCGCGCGCG 668
DB 896 -----GGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 843
QY 669 CTTCTCAGCGTGGAGGACCTTGGCTGGACACCAACACCTTGGAGCGCGCTACTGTGTA 728
DB 842 CGGCG 783
QY 729 CGCAACCTGCACACCGGTTCGTTTCAGCGATGCGCTTCAGGCGCTTGGCGGATGACGG 788
DB 782 GGNCCNNNNCCCCNNCCNNCCCCNNCCCCNNCCCCNNCCCCNNCCCCNNCCCC 723
QY 789 ACACCGCGTTCGTGCAAGTACAGCCCCCACCACCTCGTCCCGCCATCGAGACAC 848
DB 722 CCCCCGGGCCCCNNCCCCNNCCCCNNCCCCNNCCCCNNCCCCNNCCCCNNCCCC 663
QY 849 CACCGAAGACACCGCGAAGACGTACCGCGATCGGACGCTTCGCGCGCGGCGACAACGA 908
DB 662 CCCCCCCCCCCCCCCCCNNCCCCCCCCCGGGCCCCCCCCCCCCCCCCCCCC 603
QY 909 CACCGCGCGC 918
DB 602 CCCCCCNC 593

RESULT 2
AG036221
LOCUS 870 bp DNA linear GSS 01-NOV-2001

DEFINITION Pan troglodytes DNA, clone: PTB-012B17.P, genomic survey sequence.
ACCESSION AG036221
VERSION AG036221.1 GI:16563094
KEYWORDS BAC Library clone:PTB-012B17.P.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM Pan troglodytes
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 870)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
Location/Qualifiers
1..870
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-012B17.P"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 14 a 241 c 512 g 31 t 72 others
ORIGIN
Query Match 9.1%; Score 83.4; DB 17; Length 870;
Best Local Similarity 43.5%; Pred. No. 1.1e-05;
Matches 275; Conservative 0; Mismatches 356; Indels 1; Gaps 1;
QY 30 GTGGCCCGGATGGGTGTGCTGCTCGCTCTCTCTCGGTGTTCGCCCGCGGATGCA 89
DB 73 GNNCGCGCGCGCGCGCGGNTTGTGNNNNNNCGGNNNNNNCGGNNNNNN 132
QY 90 GCGGTGCGAGGAGGCTCTGGCGCGGTGGTGGACTGTCTGTGGTGGACATCTCGCGCG 149
DB 133 GGGCGCGCGCGCGGNNNGGGGGGGGGCGCNCNNGNNGGGGGGGCGNNNTGTTTGGGGGG 192
QY 150 GACGCGCGGATGCGGTGTGGAGCGCGCGCGATGTGTCCAGCTGTGCTTTCAGCGT 209
DB 193 GGGGTGTGGNNNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 252
QY 210 CATGTGTCTTTGGTGTCTCTGTGGCGTTCCTAAGGT-ATCGAACCCGACGCGGTCTTG 268
DB 253 NGGGGNGGCG 312
QY 269 GCCATTTCCAGGCGAGATCGCGCGCGCATGTGTGTGGGCGCTGAGCTTGAAGGACG 328
DB 313 NCCNCGNGGGCGCGCGCGNNNCGCGCGCGGGGGGGGGGGGGGGGGGGGG 372
QY 329 CGGCGAAGACTGTTGCGCTGCGACCGCGCGCTGGCGCGCTGTGCGCGCGCGCGCGCG 388
DB 373 CG 432
QY 389 TGGCTCAGTGCCTGCTGCTGCCCGAGGAGGTGGAGAGCTCATTCGTGAGCGGTGGCGG 448

C	1	2	3	4	5	6
	90.4	9.8	1364	14	BM810045	AGENCOURT
	83.4	9.1	870	17	AG036221	Pan trogl
	82.2	8.9	932	17	CNS0072Q	Drosophil
	80.6	8.8	1364	14	BM810045	AGENCOURT
	79	8.6	1312	14	BM914111	AGENCOURT
	78.2	8.5	691	17	AG171171	Pan trogl

THIS PAGE BLANK (USPTO)


```
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/836,821
/ CURRENT FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 32
/ LENGTH: 11220
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-836-821-32

Query Match      33.8%; Score 311; DB 9; Length 11220;
Best Local Similarity 61.3%; Pred. No. 1.9e-61;
Matches 557; Conservative 0; Mismatches 340; Indels 12; Gaps 3;

QY      4 TTCTGTTTCCCGACAGGGTGGGCGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 63
DB      6205 TTCTGTTTCCCGGACAGGGCGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 63

QY      64 TCTCCGGTGTCCCGCGGATGACAGGCGTGGAGGAGTCTGGCGCGTGGTGGAC 123
DB      6265 TCGAAGAGTTCGCGCGGCGATGGCGGAGTGGGTGCTGCTCGTACGTGGAC 6324

QY      124 TGTCTGTGTGGTACATCTCGCGCGGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 183
DB      6325 TGTCTGTGTGGTACATCTCGCGCGGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 6384

QY      184 GTGTCTCAGCTGTGCTGTTTCTGAGCTCATGTTGCTTTTGGTGTCTTGTGGCTTCTTAC 243
DB      6385 GTGTCTCAGCTGTGCTGTTTCTGAGCTCATGTTGCTTTTGGTGTCTTGTGGCTTCTTAC 243

QY      244 GGTATCGAAGCGGAGCGGCTCTTGGCCATTTCCAGGCGGAGATCGCGCGCGGATGTG 303
DB      6445 GGGGTGACCCCGCAAGCGGCTGTGCGGCCATCTCGAGGCGGAGATCGCGCGCGGATGTG 6504

QY      304 TGTGGGCGCTGAGCTCAAGAGCGCGGAGAGTGTGTGCTCGGAGCGGCGGCTG 363
DB      6505 GCGGTGCGCTGAGCTCAAGAGCGCGGAGAGTGTGTGCTCGGAGCGGCGGCTG 6564

QY      364 GCGGC---TGTGCGGCGCGGCGGAGTGGCTCAGTCCGCTGCTGCTGCCAGAGGTG 420
DB      6565 GCGGC---TGTGCGGCGCGGCGGAGTGGCTCAGTCCGCTGCTGCTGCCAGAGGTG 6624

QY      421 GAGCAGCTCATTTGTGTGAGCGGTGGGCGGCGGTTGTGGGTGGCGGCTCAACGCGCCC 480
DB      6625 GTGAGCAGCTGGGCGG-----GGTTTCAGCGGCTGTCTGCTCGCGCGGCTCAACGCGCCT 6678

RESULT 15
US-09-861-289-32
/ Sequence 32, Application US/09861289
/ Patent No. US20020110897A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/861,289
/ CURRENT FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 32
/ LENGTH: 11220
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match      33.8%; Score 311; DB 10; Length 11220;
Best Local Similarity 61.3%; Pred. No. 1.9e-61;
Matches 557; Conservative 0; Mismatches 340; Indels 12; Gaps 3;

QY      4 TTCTGTTTCCCGGACAGGGTGGGCGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 63
DB      6205 TTCTGTTTCCCGGACAGGGCGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 6264

QY      64 TCTCCGGTGTTCGCGCGGAGTGGCGGCTGGAGGAGTCTTGGCGCGTGGTGGAC 123
DB      6265 TCGAAGAGTTCGCGCGGCGATGGCGGAGTGGGTGCTGCTCGTACGTGGAC 6324

QY      124 TGTCTGTGTGGTACATCTCGCGCGGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 183
DB      6325 TGTCTGTGTGGTACATCTCGCGCGGAGTGGGCGGAGTGGGTGCTGCTCGCTCC 6384

QY      184 GTGTCTCAGCTGTGCTGTTTCTGAGCTCATGTTGCTTTTGGTGTCTTGTGGCTTCTTAC 243
DB      6385 GTGTCTCAGCTGTGCTGTTTCTGAGCTCATGTTGCTTTTGGTGTCTTGTGGCTTCTTAC 243

QY      244 GGTATCGAAGCGGAGCGGCTCTTGGCCATTTCCAGGCGGAGATCGCGCGCGGATGTG 303
DB      6445 GGGGTGACCCCGCAAGCGGCTGTGCGGCCATCTCGAGGCGGAGATCGCGCGCGGATGTG 6504

QY      304 TGTGGGCGCTGAGCTCAAGAGCGCGGAGAGTGTGTGCTCGGAGCGGCGGCTG 363
DB      6505 GCGGTGCGCTGAGCTCAAGAGCGCGGAGAGTGTGTGCTCGGAGCGGCGGCTG 6564

QY      364 GCGGC---TGTGCGGCGCGGCGGAGTGGCTCAGTCCGCTGCTGCTGCCAGAGGTG 420
DB      6565 GCGGC---TGTGCGGCGCGGCGGAGTGGCTCAGTCCGCTGCTGCTGCCAGAGGTG 6624

QY      421 GAGCAGCTCATTTGTGTGAGCGGTGGGCGGCGGTTGTGGGTGGCGGCTCAACGCGCCC 480
DB      6625 GTGAGCAGCTGGGCGG-----GGTTTCAGCGGCTGTCTGCTCGCGCGGCTCAACGCGCCT 6678
```

Qy	244	GGTATCGAACC	CGGACCGCGGTCTTGGCCATTCCAGGGCGAGATCGGGCGCGGATGTG	303
Db	6445	GGGGTACCCCGCA	AGCCGTGTGGCCACTCGAGGCGGAGATCGCGCGCGGTACGTC	6504
Qy	304	TGTGGGGCTGAGCCT	GAAGACGCGCGAAGACTGTTGGCTGCGGACGCGGCGCGCTG	363
Db	6505	GCCGGTGCCCTGAGCCT	GGAGCGCGCGCTCGTGTCTGACCCCTGCGGACGAAGTCCATC	6564
Qy	364	CCCCG---	TGTGCGGGGCGGGGGGATGCGCTCAGTGC CGCTGCTGCCACGAGGTG	420
Db	6565	GGCGCCACTCTCG	GGGCGCAGTGTCTCGCGCTGAGGAGGCGCGCGTT	6624
Qy	421	GAGCAGCTCA	TGTGTGACGGTGTGGCGGGCGGTTGTGGGTGGGCGGTCAACGGCCCC	480
Db	6625	GTGAGCGACTGGCCG	-----GGTTCGACGGGTGTCCGTGCGCGCGTCAACGGGCGCT	6678
Qy	481	CGTCTCACCGCGCTCT	GGGGGATGCGGAGCGGTGGACGAGGTGCTGGCGTACTGTGCC	540
Db	6679	ACCGCCACCGTGGT	TTTGGGCGACCCGACCCAGATCCAAAGAGCTCGCTCAGGCGTGTGAG	6738
Qy	541	GGCACCGGGTGGCGG	CCCGCGGATCCCGGTTCGACTATGCTTCGCACTGCCCTCCCATGTG	600
Db	6739	GCCGACGGGGTCCG	CGCACGATCATCCCGTTCGACTACGGCTCCCAACGCGCCACGTC	6798
Qy	601	CAGCCCCCTGCGGG	AGGAGTGTCTGGAGTCTGCTGGGGACATCAGCCGCGAGCGCTCCGGC	660
Db	6799	GAGACCATCGAGCG	AACTCGCCGACGTCTCTGGCGGGTTGTCCCTCCGACACCCGAG	6858
Qy	661	GTGCGGTTCTTCTCC	ACGGTGGAGGGCACCTGGCTGGACACCAACCCCTGGACGCGCC	720
Db	6859	GTCCCCCTTCTTCC	ACCTCGAAGCGCCTGGATCACCGAACC CGCCCTCGACGCGCGC	6918
Qy	721	TACTGGTACCGCA	ACTGCAACCGGTTCGTTTCAAGGATGCGCTCGAGGCCCTCG--	778
Db	6919	TACTGGTACCGCAA	CCCTCGCATCTGTGTGGGCTTCGCGCGCGCTCGAAACCTCGGCC	6978
Qy	779	-CGGATCAGCGACA	CCCGCTTCTCGTTCGAGTCAAGCCCGCACCCACCTCTCGTCCCGCC	837
Db	6979	ACCGAAGAGCTTCA	CCCACTTCTGTGAGGTCAAGCGCCACCCCGCTCTCAACGCGC	7038
Qy	838	ATCGAAGACACCA	CCCGGAAGACCTCACCGGATCGGCAAGCTCTCCGCGC	897
Db	7039	CTGCCCCGAGAC	CGTCAACCGGACTCGGCACCTTCGCGCTGACACGGCGACGACCGC	7098
Qy	898	GGCGACCAAC	906	
Db	7099	CTCACCACC	7107	

RESULT 13

US-09-988-384B-32

; Sequence 32, Application US/09988384B

; Publication No. US20030073824A1

GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

APPLICANT: Zhao, L

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600.536US1

;	CURRENT APPLICATION NUMBER:	US/09/988,384B
;	CURRENT FILING DATE:	2001.11.10

; CURRENT FILING DATE: 2001-11-19
 ; PRIOR APPLICATION NUMBER: PCT/US

; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1998-05-25

;
:
:

; PRIOR APPLICATION NUMBER: US 09/105,537
 : PRIOR FILING DATE: 1998-06-26

; FRICK FILING DATE: 1998-08-28
: NUMBER OF SEQ ID NOS: 53; NUMBER OF SE
; SEO ID NO 32

; LENGTH: 112

```

;
; TYPE: DNA

```

ORGANISM:

US-09-988-384B-32

Query Match	33.8%;	Score 311;	DB 9;	Length 11220;
Best Local Similarity	61.3%;	Pred. No. 1.9e-61;		
Matches 557; Conservative	0;	Mismatches 340;	Indels 12;	Gaps 3;

RESULT 14

US-09-836-821-32

; Sequence 32, Application US/09836821

Publication No. US20030087405A1

```
Db 5645 CTCCCGGACC 5656 |||
RESULT 11
US-09-793-708-19
; Publication No. US09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 34.5%; Score 318; DB 9; Length 38506;
Best Local Similarity 63.4%; Pred. No. 4.3e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;

QY 4 TTCTGTTTCCCGACAGGTTGGCAGTGGCGGGATGGGTGTGGTCTGCTCGCCTCC 63
Db 1783 TTCTGTTTCCCGACAGGTTGGCAGTGGCGGGATGGGTGTGGTCTGCTCGCCTCC 63
QY 64 TCTCCGTTTCCCGACAGGTTGGCAGTGGCGGGATGGGTGTGGTCTGCTCGCCTCC 123
Db 1843 TCCGCGTTTCCCGACAGGTTGGCAGTGGCGGGATGGGTGTGGTCTGCTCGCCTCC 1902
QY 124 TGTCTGTGTGGACATCTGCGCGGAGCGGGGATGGGTGTGGGAGCGGGCGAT 183
Db 1903 TGTCTGTGTGGACATCTGCGCGGAGCGGGGATGGGTGTGGGAGCGGGCGAT 1962
QY 184 GTGGTCCAGCTGTGCTTTACAGCTGATGGTGTGTTTGGTGTCTGTGGCTTCTTAC 243
Db 1963 GTCTGTGACCTGTGCTTTACAGCTGATGGTGTGTTTGGTGTCTGTGGCTTCTTAC 243
QY 244 GGTATCGAATCCGACCGGTTCTTGGCCATTCAGGGGAGATCGCGCGCGGATGTG 303
Db 2023 GGGGTGACCCCGACCGGTTCTGCGCCATTCAGGGGAGATCGCGCGCGGATGTG 2082
QY 304 TGTGGGCGCTGAGCTGAAGACCGCGGAGACTGTTGCGTGGCGAGCGGGCGGTG 363
Db 2083 GCGGTGCTTGTGCTGTGACGAGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 2142
QY 364 GCGGC---TGTGGGCGCGCGGCGGCTTGTGGCTTGTGGCTTGTGGCTTGTGGCTTGT 420
Db 2143 GCGGCGCACTGCGCGGCAAGGCGGCTGCTGTGCTTGTGGCTTGTGGCTTGTGGCTTGT 2202
QY 421 GAGCAGCTATTGTTGAGCGGTGGGCGGCGGCTTGTGGCTTGTGGCTTGTGGCTTGT 480
Db 2203 CTGGAGCGACTGGCCG-----GGTTCGACGGGCTGTCCGTGCGCGCTGTGAACGGGCCC 2256

QY 481 CGTCCACCGCGCTCTCGGGGGATGCCGAGCGGTGGAGAGGTCTCTGGCGTACTGTGCC 540
Db 2257 ACCGCCACCGTGTCTCTCGGTGACCCCGTACAGATCGAAGAGCTTGTCTGGCGGTGTGAG 2316
QY 541 GGCACCGGGTGGCGGGCCCGCGGATCCGGTTCGATATGCTTCGACTGTGCCCGCATGTG 600
Db 2317 GCCGATGGGTCCGTCGCGGGTCTATTCCTCGACTACGGGTCCACAGCGCGGAGTTC 2376
QY 601 CAGCCCTGCGGAGAGGTTCGTGAGCTGTGGGGACATCAGCCCGACCGTCCGCG 660
Db 2377 GAGATCATCGAGAGGAGCTCGCGAGGTCTCTGCGCGGGCTCAGCCCGAGCTCCGCGC 2436
QY 661 GTCCCGTTCTTCTCCACCGTGGAGGACCTTGGTTCGACACACCAACCTTGGACCGCGC 720
Db 2437 GTGCGTTCTTCTCGACACTCGAAGCGCGCTGGATCACGAGCGCGTCTCGACCGCGC 2496
QY 721 TACTGTTACCGAACCTTCGACCGCGTTCGTTTACGCGATGCGTTCAGGCGCTCG-- 778
Db 2497 TACTGTTACCGAACCTTCGACCGCGTTCGTTTACGCGATGCGTTCAGGCGCTCG-- 2556
QY 779 -CGGATGAGGACACCGCGTCTTCTGCGAAGTTCAGCGCCCGACCCCTCTGTCGCCGCC 837
Db 2557 ACCGACGAGGCTTCCACCACTTCGTGAGGTTCAGCGCCCGACCCCTCTCACCATGGCC 2616
QY 838 ATCGAAGACACC 849
Db 2617 CTCCCGCGGACC 2628

RESULT 12
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 33.8%; Score 311; DB 9; Length 11220;
Best Local Similarity 61.3%; Pred. No. 1.9e-61;
Matches 557; Conservative 0; Mismatches 340; Indels 12; Gaps 3;

QY 4 TTCTGTTTCCCGACAGGTTGGCAGTGGCGGGATGGGTGTGGTCTGCTCGCCTCC 63
Db 6205 TTCTGTTTCCCGGAGACAGGACACGAGTGGCGCGGGATGGGTGTGGTCTGCTCGACGTG 6264
QY 64 TCTCCGTTTCCCGCGCGGATTCAGGCGGTTCAGGAGAGGTCTCTGGCGCGGTGGAGT 123
Db 6265 TCGAAGGAGTTCGCGCGCGGATTCAGGCGGTTCAGGAGAGGTCTCTGGCGCGGTGGAGT 6324
QY 124 TGTCTGTGGTGGACATCTCTCGCGCGGATTCAGGCGGTTCAGGAGAGGTTCAGGAGT 183
Db 6325 TGTCTGTGGTGGAGTTCAGGCGGTTCAGGCGGTTCAGGAGAGGTTCAGGAGT 6384
QY 184 GTGTCTGAGCTGCTCTTTCAGGCTGATGTTTGGTGTCTTTCAGGCTGCTTTCAGGCT 243
Db 6385 GTCTGAGCTGCTGCTTTCAGGCTGATGTTTGGTGTCTTTCAGGCTGCTTTCAGGCT 6444
```



```
RESULT 6
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match      34.5%; Score 318; DB 10; Length 13842;
Best Local Similarity 63.4%; Pred. No. 4.9e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;

QY      4 TTCTGTTTCCCGACAGAGGTGGCGAGTGGCGGGAGTGGGTGCTGCTCGCTCC 63
Db      1900 TTCTGTTTCCCGACAGAGGTGGCGAGTGGCGGGAGTGGGTGCTGCTCGCTCC 1959

QY      64 TCTCCGCTGTCGCGCGGGATGACGGGTGCGAGGAGGCTCTGGCGCGGTGGGAC 123
Db      1960 TCCGCGGTGTTCCGCGCGGCATGCGCGAATGCGAGCGCGCACTCTCCCGTACGTCGAT 2019

QY      124 TGGTCTGTGGTGCATCTCGCGCGGACCGCGGGGATGCGGTGTTGGAGCGGCGCAT 183
Db      2020 TGGTCTGTGGTGCATCTCGCGCGGACCGCGGGGATGCGGTGTTGGAGCGGCGCAT 2079

QY      184 GTGTTCACGCTGTGTTGAGTGCATGTTGCTTTGGTGTCTGTTGGTGTCTGTTGGGTTCTTAC 243
Db      2080 GTGTTCACGCTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2139

QY      244 CGTATCGAACCCGACGCGCTCTTGGCCATTCCAGGGGAGATCGCGCGCGCGCATGTG 303
Db      2140 GGGGTGACGCGCCAGCGCGTCTGCGCCACTCGAGGGGAGATCGCGCGCGCGCATGTG 2199

QY      304 TGTGGGCGCTGAGCTGAAGGACCGCGGAGACTGTTGCGCTGCGCAGCGCGCGCTG 363
Db      2200 GCGGTGCGCTGAGCTGAGACGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2259

QY      364 GCGCG---TGTGGGGGCGGGGCGGATGGCTCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      2260 GCGCGCCCACTCGCGCGGCAAGGGCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2319

QY      421 GAGCAGCTCATGTGTGAGCGGTGGCGGGGCGGTTGTTGGTGGCGCGGTGCAACGCGCCC 480
Db      2320 CTGGAGCGACTGGCGG-----GGTTGACGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2373

QY      481 CGCTCCACCGCGCTCTCGGGGGATGCCAGGGCGGTGACAGAGGTGCTGGCGTACTGTGTC 540
Db      2374 ACCGCCACCGTGTCTCCGGTGACCCCGTACAGATCGAAGACTTGTCTGGGCGTGTGAG 2433

QY      541 GGCACCGGGGTGCGGGCGCGGAGATCCCGGTGCACTAGCTCTGCACTGCTGCTGCTGCTGCTGCTGCTG 600
Db      2434 GCGGATGGGGTCTCGTGGCGGTCTATTCGCTGCACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2493

QY      601 CAGCCCTGCGGAGAGTGTGCTGAGCTGTGCTGGGACATCAGCCCGCGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db      2494 GAGATATGAGAGGAGCTGCCAGGTCTCTCGCGGGCTCAGCCCGGCTCAGCCCGGAGCTCCGCGC 2553

QY      661 GTGCGGTTCTTCTCCACGTTGAGGGCACCTGGCTGGACACCAACACCTTGACGCGCGCC 720
```

```
Db      2554 GTGCGGTTCTTCTCGACACTCGAAGCGCGCTGGATCACGAGCGCGTGTCTGACGCGCGC 2613
QY      721 TACTGTATCCGCAACCTCACCAGCGGTCCGTTTTCAGGATGCGGTCCAGGCGCTGG--- 778
Db      2614 TACTGTATCCGCAACCTCGGCCATCGTGTGGCTTTGCGCGCGCGCTGAGACCTTGGCC 2673
QY      779 -CGGATGACGACACCGCGTCTTTCGTTGAAAGTACAGCCCCCACCACCTCGTCCCGCGC 837
Db      2674 ACCGAGAGGCTTCAACCACCTTCTGTCAGGTACGAGCGCCACCCCGTCTCTACCATGGCC 2733
QY      838 ATCGAAGACACC 849
Db      2734 CTCCCGGGGACC 2745

RESULT 7
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match      34.5%; Score 318; DB 9; Length 36778;
Best Local Similarity 63.4%; Pred. No. 4.4e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;

QY      4 TTCTGTTTCCCGACAGAGGTGGCGAGTGGCGGGATGGGTGTCGCTGCTCGCTCC 63
Db      3641 TTCTGTTTCCCGAGGAGGCGACGAGTGGCGCGCATGGGTGCGAACTGCTGGACTCT 3700

QY      64 TCTCCGCTGTTTCGCGCGCGGATGACGCGTGCAGAGAGGCTCTTGGCGCGTGGGTGGAC 123
Db      3701 TCCGCGGTGTTTCGCGCGCGCATGCGCGAATGCGAGGCGCGCACTCTCCCGTACGTCGAC 3760

QY      124 TGGTCTGTGTGGACATCTCTCGCGCGGGAACGCGGGGATGCGGTGTTGGAGCGGCGCGAT 183
Db      3761 TGGTCTGTGTGGAGCGCTGACGCGAGCGCGCGCGCGCGCGCGCTGCGAGCGGTGCGAT 3820

QY      184 GTGTTCAGCGCTGCTGTTTCAGGTGCTATGTTGCTTTTGGTGTCTGTTGGCTTCTTAC 243
Db      3821 GTGTGACGCTGTGAGCTTCGCGCTCATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3880

QY      244 GGTATCGAACCCGACGCGGTCTTTGGCCATTTCACAGGCGGAGATCGGGCGCGCGCATGTG 303
Db      3881 GGGGTGACGCGCCAGGCGGTCTGTCGCGCACTCGCAGGCGGAGATCGCGCGCGCTGACGTC 3940

QY      304 TGTGGGCGCTGAGCTCGAAGGACCGCGCGAAGACTGTTTGGCTGCTGCGAGCGCGCGCGCTG 363
Db      3941 GCGGTCGCTGAGCTGAGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4000

QY      364 GCGCG---TGTGGGCGCGGGCGGATGSCCTCAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db      4001 GCGCGCCACCTCGCGCGGCAAGGGCGGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4060

QY      421 GAGCAGCTCATTTGTTGAGCGGTGGCGGGCGGTGTTGGGTGGCGGGGCTCAACGCGCGCC 480
```

```
Db 1900 TTCTGTTTCCCGGCGAGGCGACGCAAGTGGGCGGCGCATGGGTGCCGAATCTGTGACTCT 1959
Qy 64 TCTCCGGTCTTCGCGCGGCGATGCAAGCGTGCAGGAGGCTCTGCGCGCGTGGTGAC 123
Db 1960 TCGCGGTGTTTCGCGCGGCGCATGGCGAATCGAGGCGCGACTCTCCCGTACGTGCAC 2019
Qy 124 TGGTCTGTGGTGACATCTGCGCGGGAATCGCGGGGATGCGGTGTGGAGCGGCGCAT 183
Db 2020 TGGTCTGTGGGCGGTCTGACGCGGCGGCGCGGCGGCGCGCGGCGGCGCAT 2079
Qy 184 GTGGTCCAGCCTGTGCTTGTGACGCTCATGGTGTCTTTGGCTGTCTTGGGCTTCTTAC 243
Db 2080 GTCTGTGACGCTGTGACGCTCATGGTGTCTTGGGCTGTCTTGGGCTTCTTAC 2139
Qy 244 GGTATCGAACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
Db 2140 GGGGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2199
Qy 304 TGTGGGCGCTGAGCTGAAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363
Db 2200 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2259
Qy 364 GCGCG---TGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 2260 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2319
Qy 421 GAGCAGCTCATTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 2554 GTGCGGTTCTTCTCGACACTCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2613
Qy 721 TACTGTACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778
Db 2614 TACTGTACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2673
Qy 779 -CGGATGACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 837
Db 2674 ACCGACGAGGCTTACCCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2733
Qy 838 ATCGAAGACACC 849
Db 2734 CTCCCCGGGACC 2745
```

RESULT 5

```
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836, 821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
```

```
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

Query Match 34.5%; Score 318; DB 9; Length 13842;
Best Local Similarity 63.4%; Pred. No. 4.9e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;

Qy 4 TTCTGTTTCCCGGCGAGGCGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63
Db 1900 TTCTGTTTCCCGGCGAGGCGACGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1959
Qy 64 TCTCCGGTCTTCGCGCGGCGATGCAAGCGTGCAGGAGGCTCTGCGCGCGTGGTGAC 123
Db 1960 TCGCGGTGTTTCGCGCGGCGCATGGCGAATGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 2019
Qy 124 TGGTCTGTGGTGACATCTTGTGACGCTCATGGTGTCTTTGGCTGTCTTGGGCTTCTTAC 183
Db 2020 TGGTCTGTGGGCGGCGTGTGACGCTCATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2079
Qy 184 GTGGTCCAGCCTGTGCTTGTGACGCTCATGGTGTCTTTGGCTGTCTTGGGCTTCTTAC 243
Db 2080 GTCTGTGACGCTGTGACGCTCATGGTGTCTTGGGCTGTCTTGGGCTTCTTAC 2139
Qy 244 GGTATCGAACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
Db 2140 GGGGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2199
Qy 304 TGTGGGCGCTGAGCTGAAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363
Db 2200 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2259
Qy 364 GCGCG---TGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 2260 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2319
Qy 421 GAGCAGCTCATTTGTGAGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 2320 CTGGAGCGACTGGCGG-----GGTTCGACGCGGCTGTCCGCGCGGCTGTGAAACGCGGCGG 2373
Qy 481 CGCTCCAGCGCGCTCTCGGGGATGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 2374 ACCGCGACCGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGTCTCGGCGGTGTGAG 2433
Qy 541 GGCACCGGGGTGCGGGCGCGCGGATCCCGGTTCGACTATGCTCGCACTTGCCTCCCGCGGCGG 600
Db 2434 GCGGATGGGTCCGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2493
Qy 601 CAGCGCGGCGGAGGAGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 2494 GAGATCATGAGAGCGAGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2553
Qy 661 GTGCGGTTCTTCTCGACACTCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 2554 GTGCGGTTCTTCTCGACACTCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2613
Qy 721 TACTGTACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778
Db 2614 TACTGTACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2673
Qy 779 -CGGATGACGCGCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 837
Db 2674 ACCGACGAGGCTTACCCACTTCTTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2733
Qy 838 ATCGAAGACACC 849
Db 2734 CTCCCCGGGACC 2745
```

[illegible]

```

RESULT 3
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

```

Db	2140	GGGGTGACGCCCCAGCGCGTCTGTCGGCCACTCGCAGGGCGAGATCGCCGCGCGGTACGTC	2199
Qy	304	TGTGGGCGCTGAGCCTGAAGAGACGCGCGCGAAGACTGTTGCGCTGCGCAGCGCGGCGCTG	363
Db	2200	GCGGTGCCTGAGCCTGGACGACGCCGCTCGTCTGCTGACCCCTGCGCAGCAAGTCCATC	2259
Qy	364	CCGC-----TGTGCGGGCCGGGCGGCGATGGCCTCAGTCGCGCTGCTGCCCAGGAGGTG	420
Db	2260	GCCGCCACCTCGCGCGCAAGGGCGGCGATGTGTCTCCCTCGCGCTGACGAGGACGCCGTC	2319
Qy	421	GAGCAGCTCATGTGTGAGCGTGTGGCGGGCGGTTGTGGTGGCGGCGGTCACCGGCCCC	480
Db	2320	CTGAGCGACTGGCG-----GTTTCGACGGGCTGTCCGTGCGCCGTGTGAACGGGCC	2373
Qy	481	CGTCTCACCGCGCTCTCGGGGGATGCCGAGCGGTTGACGAGGTGCTGGCGCTACTGTGCC	540
Db	2374	ACGCCACCGTGGTCTCCGGTGAACCCGTTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG	2433
Qy	541	GGCACCGGGTGGGGCGCGCGGATCCCGTTCGACTATGCTTCGACTGCCCCCATGTG	600
Db	2434	GCAGATGGGTTCGCTGCGCGGGTCATTTCCTCGTTCGACTACGCTGCCACAGCGCGCAGTTC	2493
Qy	601	CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCGCGTCCGGC	660
Db	2494	GAGATCATCAGAGCGAGCTCGCGGAGTCTTCGCGGGCTCAGCCCGCAGGCTCGCGC	2553
Qy	661	GTCCGCTTCTTCTCACGTTGGAGGGCACTGCTGGACACCAACCTTCGAGCGCGCC	720
Db	2554	GTGCGCTTCTTCGACACTCGAAGGGCGCTGGATCACAGGCGCGTGTCTCAGCGCGCGC	2613
Qy	721	TACTGGTACCGCAACTGACACCGCGTTCGTTTTCAGCGATCGCGTTCAGGCGCCCTGG--	778
Db	2614	TACTGGTACCGCAACTGCGCCCATCGTGTGGGCTTGGCCCCGCGCGTCGAGACCTTGCC	2673
Qy	779	-CGGATCAGCGACACCGCGTCTTCGTGGAAGTCAGCCCCCACCCACCTCTGTCGCCGCC	837
Db	2674	ACCGAGGGCTTACCCACTTCTGTCGAGTTCAGCGCCACCCCGTCTCTCACCATGCC	2733
Qy	838	ATCGAAGACACC	849
Db	2734	CTCCCCGGACC	2745

```

RESULT 4
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-25
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

```

Qy 4 TTCTGTGTTTCCCGGACAGGTGGGCACTGGGCGGGATGGGTGTCGCTCTGCTCGCCTCC 63

Db 24765 GGTGGTCTGCTGAGGTGTTGCGC---GGTGGTGGTGAAGCTGTTCTTGGGGGGGTT 24821
Qy 181 GATGGTGTACAGCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTGTGGGGTTC 240
Db 24822 GATGGTGTACAGCGCGCTGTGGGGCGGTGATGGTGTCTACTGGCCGGACCTGGCGGTAT 24881
Qy 241 TACGGTATCAAAACCCAGACCGGTCTTTGGCCATTTCCAGGGCGAGATCGCGGCCGCGCAT 300
Db 24882 TACGGTGTAGAGCTGCTGCGGTGTGGGGCATTCGACGGGTGAGATCGCTGGCGTTGT 24941
Qy 301 GTGTGTGGGGCGCTGAGCTGAAGGACCGCGCGAGACTGTTGCGCTGCGCAGCGCGGG 360
Db 24942 GTGGCTGGGGGGTTGAGTCTGGCCGATGGTGGCGGGTGTGTGTGGGAGCGCGGG 25001
Qy 361 CTGGCCGCTGTGGGGCGGGCGGCATGGCTCAGTCCGCTGCTGCCCTGCCAGGAGTG 420
Db 25002 ATCGCCCGGATCGCTGGTGGGGCGGCATGGTCTGGTGGTCTTTTCAGC---TGAAGCT 25058
Qy 421 GAGCAGCTCATTTGGTGAAGCGGTGGGGCGCGTGTGGGTGGCGCGGTCAACGCCGCC 480
Db 25059 GTCCGACCATGCTCGACACCTACGGTGGCGGGTTTCGGTGGCGGGTCAATGGCGCG 25118
Qy 481 CGCTCACCGCGCTCTCGGGGAGTCCGAGGCGGTGGAGAGGTGCTGGCGTACTGTGCC 540
Db 25119 TCCTCAGCGCTGTGTCCGGTGAAGTCCAGGCGCTGGATGAGTTGTTGGCGCGTGTGAG 25178
Qy 541 GGCACCGGGTGGCGGGCGGGGATCCCGGTGAGTATGCTCGCACTGCCACTGCCCGCATG 600
Db 25179 CGGAGGGTGTCCGGGCTGTGTGTCCGGTGGACTATGCTCCCACTCCGCGCAGATG 25238
Qy 601 CAGCCCTCGGAGAGGTGTTGGAGCTGCTGGGGGACATCAGCCCGCAGCGCTCCGCG 660
Db 25239 GACCACTTACGGGATGAGCTGTGGAAGCGTGGCGGACATCACTCCGCAACATTCAGT 25298
Qy 661 GTCCCGTTCTTCTCAGGTGGAGGCACTGTGGTGGACACACCAACCTTGACGCGGCC 720
Db 25299 GTTCGGTTCTTCGACGGTGAAGCGGAGTGGCTGGACACACCGCTCTGGATGGGG 25358
Qy 721 TACTGTGTACCGAACCTGCACAGCGGTCCGTTTCAGCGATGCGTCCAGGCGCTGGCG 780
Db 25359 TACTGTGTACGANTCTGGGAGACGGTCCGGTTCAGAGAGCGCTCGAAGGGCTCGTG 25418
Qy 781 GATGACGACACCGCGTCTTCTGCGAAGTCAAGCCCGCCACCCACCTCGTCCCGGCATC 840
Db 25419 GCTCAGGGATGGCGGCTTGTGCGAGTGCAGCCCGCACCCCGTCTGCTCCCGGTATC 25478
Qy 841 GAAGACACCAAGACACCGCGGAGAGGTCAACCGGATCGGAGCGCTCGCGCGCGCG 900
Db 25479 GAGCAGACCTCGACGCGCTCGACCAAGAGCGCGCGTACTCGGCTCCCTGGCGGTGAC 25538
Qy 901 GACAACGACACCGCGCGCTTC 921
Db 25539 GAAGCGGGCTGACCGCACTC 25559

RESULT 2

US-09-735-056-29
; Sequence 29, Application US/09735056
; Publication No. US20030013662A1
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Staehli, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaolan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.

; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/735,056
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,003
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuso
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 4952.US.P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847)-938-3137
; TELEFAX: (847)-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-735-056-29

Query Match 39.4%; Score 362.8; DB 9; Length 1010;
Best Local Similarity 63.7%; Pred. No. 6.1e-73;
Matches 585; Conservative 0; Mismatches 327; Indels 6; Gaps 2;
Qy 1 GTCTTCTGTTTCCCGACAGGTGGGAGTGGCGCGGGATGGGTGCTGCTCGCC 60
Db 13 GTGTCTGTTTCCCGCGCAGGGTTCGAGTGGCGCGGATGGCGCGGGTGTCTGGAG 72
Qy 61 TCCTCTCGGTGTTCCCGCGCGGATGAGCGGTGCGAGGAGCTCTGGCGCGGGTG 120
Db 73 CGGTCCGGCGGTTCGAGTGGCGCGCTCTGCGAGCGCGCGCTCGCGCGTACCTC 132
Qy 121 GACTGCTGTGTGGACATCTGCGCGCGGACCGCGGGGATGCGGTGTGGAGCGGGCC 180
Db 133 GGCTGCTGTGTGAGGTGCTGCGCGGGGACCGGACCGCGCTCTGCTCAACCGGTC 192
Qy 181 GATGTGTCAGCTGTGCTGTTCAAGCTCATGTGTGTTTGGTGTCTGTGGGTTC 240
Db 193 GACGTCTGACGCGGTGCTGTTACGATGATGTTCTGCTCGCGCGGTCTGGGTGCG 252
Qy 241 TACGTTATGAAACCGCGCGGTCTTGGCCATTCGAGGGGAGATCGCGCGCGCAT 300
Db 253 CTGGGGGTGGAACCGCGCGCGGTCTGCGGCGCATCTGCGAGGGTGAATCGCGCTGCCCAT 312
Qy 301 GTGTGTGGGGCTGAGCTGAAGGACGCGCGGAAGACTGTTTGGCTGCGCAGCGGGCG 360
Db 313 GTGCGCGGTGGCTGTCTGCTGAGACACTGCGCGCGGATCGTCCCTGCGCAGTGGGG 372
Qy 361 CTGGCGCTGTGGCGGGCGCGCATGGCTCATGTGCGCTGCTGCTGCGCGAGGTG 420
Db 373 TGGCTCGGACTGGCGGGCAAGGGCGCATGTGTGGCGGTGCCGATGCGCGCGAGGTG 432
Qy 421 GAGCAGCTCATTTGTTGAGCGGTGGCGGGGGTGTGGTGGCGCGGTCAACGGCCCC 480
Db 433 CGGCGCGCGCT---CGTACGCTGGGGGACCGTCTGGCGGTGCGCGCGTCAACGGCCCC 489
Qy 481 CGCTCACCGCGCTCTCGGGGATCCGAGGCGGTGGAGGAGTGTGCTGCGTACTGTGCG 540
Db 490 GGTTCCTGCGCGCTGCGAGGCGACCGGAGGCGCTGGCGGAACTGTTGGCGTGTGACC 549
Qy 541 GGCACCGGGTGGCGGGCGCGCGGATGCC---GGTGGACTATGCTCTGCACTGCGCCCAT 597

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 108.842 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgttcccgagca.....acaagacaccgcgcgttc 921

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues
Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	373.8	40.6	50937	9	US-09-808-880-1
2	362.8	39.4	1010	9	US-09-735-056-29
3	318	34.5	13842	9	US-09-860-846-30
4	318	34.5	13842	9	US-09-988-3848-30
5	318	34.5	13842	9	US-09-836-821-30
6	318	34.5	13842	10	US-09-861-289-30
7	318	34.5	36778	9	US-09-860-846-5
8	318	34.5	36778	9	US-09-836-821-5
9	318	34.5	36778	10	US-09-861-289-5
10	318	34.5	37948	9	US-09-988-3848-5
11	318	34.5	38506	9	US-09-793-708-19
12	311	33.8	11220	9	US-09-860-846-32
13	311	33.8	11220	9	US-09-988-3848-32
14	311	33.8	11220	9	US-09-836-821-32
15	311	33.8	11220	10	US-09-861-289-32
16	308.6	33.5	1035	9	US-09-735-056-30
17	305.8	33.2	4689	9	US-09-860-846-34
18	305.8	33.2	4689	9	US-09-988-3848-34
19	305.8	33.2	4689	9	US-09-836-821-34

20	305.8	33.2	4689	10	US-09-861-289-34	Sequence 34, Appl
21	298.4	32.4	68750	9	US-10-014-717-1	Sequence 1, Appl
22	280.6	30.5	4041	9	US-09-860-846-36	Sequence 36, Appl
23	280.6	30.5	4041	9	US-09-988-3848-36	Sequence 36, Appl
24	280.6	30.5	4041	9	US-09-836-821-36	Sequence 36, Appl
25	280.6	30.5	4041	10	US-09-861-289-36	Sequence 36, Appl
26	187.4	20.3	15872	9	US-09-860-846-1	Sequence 1, Appl
27	187.4	20.3	15872	9	US-09-988-3848-1	Sequence 1, Appl
28	187.4	20.3	15872	9	US-09-836-821-1	Sequence 1, Appl
29	187.4	20.3	15872	10	US-09-861-289-1	Sequence 1, Appl
30	164.4	17.9	5484	9	US-09-712-363-115	Sequence 115, App
31	152.8	16.6	1030	9	US-09-735-056-2	Sequence 2, Appl
32	147.4	16.0	4851	9	US-09-712-363-116	Sequence 116, App
33	147	16.0	4209	9	US-09-712-363-20	Sequence 20, Appl
34	144.6	15.7	426	9	US-10-125-815-5	Sequence 5, Appl
35	138.2	15.0	426	9	US-10-125-815-6	Sequence 6, Appl
36	64.4	7.0	3331	9	US-09-373-658-31	Sequence 31, Appl
37	62.2	6.8	925	9	US-09-735-056-1	Sequence 1, Appl
38	61.8	6.7	12733	9	US-10-032-393-47	Sequence 47, Appl
39	61.8	6.7	12739	9	US-10-032-393-8	Sequence 8, Appl
40	60.4	6.6	88421	9	US-09-976-059-1	Sequence 1, Appl
41	58.6	6.4	3372	9	US-10-067-457-2	Sequence 2, Appl
42	58.6	6.4	38186	9	US-09-373-658-38	Sequence 38, Appl
43	57.8	6.3	2307	9	US-09-893-519A-87	Sequence 87, Appl
44	57.2	6.2	4257	9	US-09-825-288A-1	Sequence 1, Appl
45	57	6.2	520	9	US-10-184-644-332	Sequence 332, App

ALIGNMENTS

RESULT 1
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-808-880-1
Query Match 40.6%; Score 373.8; DB 9; Length 50937;
Best Local Similarity 64.4%; Pred. No. 1.2e-75;
Matches 593; Conservative 0; Mismatches 322; Indels 6; Gaps 2;
QY 1 GTCTTCGTTTCCCGGACAGGTGGGCGAGTGGGCGCGGATGGGTGCTGCTCGCC 60
Db 24645 GTGTTTCGTTTCTCTGTCAGGGTTCGAGTGGGTGGGATGGCGCTGGTGTGGAT 24704
QY 61 TCCTTCGTTTTCGCGCCCGCGGATGAGCGTGGCGGATGAGCGTGGCGGCTGGGTGG 120
Db 24705 GCGTTCGTTTTCGCGGAGCGGTGGCGGATGAGCGTGGCGGCTGGGTGGATCGGTGACG 24764
QY 121 GACTGCTGTGTTGGACATCTCTGCGCGGACGCGGGGATGCGGTGTGGGAGCGGGCC 180


```
Db 21250 GCACGGCGCCT---CCAGCGCTATGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306
QY 481 CGGTCCACCGCGCTCTCGGGGGAATGCCAGGCGGCTGAGAGAGTGTGAGCTATGTGCC 540
Db 21307 CGTTTACGACGATCTCCGGCGAGCCCTCGCTGCGCGCCCTGCTCCGCGATCTGAG 21366
QY 541 GGCACCGGGGTGCGGCGCGCGCGATCCGAGTCCGATATGCCCTCGCATGCCCCCATGTG 600
Db 21367 TCCGAGGGGCGTCTTCCCTCCAGCTGATTAAGACTTCGCCCTCCCATCTCCGCGCAGT 21426
QY 601 CAGCCCTTGGGAGGAGTTGCTGAGCTGTGCGGGAACATCAGCCCGCAGCCGTCGGC 660
Db 21427 GAGTCGATTGCGAGAGCTCTCGATCTCTGTGCTGAGCCGCGCTGACGCGG 21486
QY 661 GTGCCCTTCTTCTCCACGATGAGGGACCTGTGCTGAGACCAACCTTGAACGCGGC 720
Db 21487 GTCCCGTTCTTCTCCACGATGAGGGACCGCGCGATGACCGGAGCGAGCTCGAGCGCGC 21546
QY 721 TACTGTTACCGCAACCTGACACGAGCGATCCGTTTACGCGATGCCCTCCAGGCCCTGCG 780
Db 21547 TACTGTTACCGCAACCTGAGCGAGCGGCTCGCTTGCAGACGCTGTGCAAGGCTTCCT 21606
QY 781 GATGACGACACCGCGTCTTGTGTAAGTCAGCCCAACCCCACTCTGTCGCCGATC 840
Db 21607 GCCGAGAAACATGCTTCTTGTGAGGTAGGCCCAAGTCTGTGCTGACCTTGAGCCTTG 21666
QY 841 GAAGACACCAACCGAAGACACCGCGGAAGAGTCAACCGCATTCGGCAGCCTCCGCGCGGC 900
Db 21667 CAGGAGCTCTCGAAGCGTGCAGCGCTCGGCGCGGTGTGCTCTCTGTGAGCGAC 21726
QY 901 GACACGACACCGCGCGCTTC 921
Db 21727 GAAGGGATCTACGCGCTTC 21747
```

Search completed: June 18, 2003, 01:06:54
Job time : 51.8967 sec

Db 20950 GACTGATCGCTGTCGCGGTGCTCCCGGAGAGAGCGCGCCCGCTCGACCGAGT 21009
Qy 181 GATGATGTCAGCCTGTGCTGTTTCAAGCTCATGATGTTTGTGCTGCTGTGAGCTTCC 240
Db 21010 GACGTGTCAGCGCGCGGCTTCTGATGATGATGCTCGCTGCGCCCGCTGTGAGCTCC 21069
Qy 241 TACGATATGAACCCGACCGCGCTCTTGGCCATTCCGAGGCGAGATCGCGCGCGCAT 300
Db 21070 ATGGGCGGTGAGCGCCGACCGGCGGTGCGCATAGCGAGGCGAGATCGCGCGCGCTGT 21129
Qy 301 GTGTGTGGGGCGGTGAGCGCTGAAAGACGCGGCGAAGACTTGTGCGCTGCGGACCGCGG 360
Db 21130 GTGGCGGGCGCGGTGTGCTGCGTGAAGCGCTGCGCAACCTGTGCGCTGTGCGACCGT 21189
Qy 361 CTGGCGCTGTGCGGCGCGCGCGCGCATGAGCTCAATGCGCGCTGCTGCGCCGAGAGTG 420
Db 21190 CTGTGTGAGCTCGCCCGCGGCGGCGGCGCATGCGCGCGGTGAGCTGCGGAGCGCGAGT 21249
Qy 421 GAGCAGCTCATTTGTGAGCGGTGAGGCGGCGCGGCGGTGTGAGGCGCGGCTCAACGCG 480
Db 21250 GACGCGCGCGCT---CGAGCGCTATGAGCGATCGGCTCTCCATCGGCGCGATCAACGCG 21306
Qy 481 CGCTCCACCGCGCTCTCGGCGGATGCGGCGGTGAGAGAGTGTGTGCGATCTGTGCG 540
Db 21307 CGTTTACAGCATCTCGCGCGAGCGCCCTGCGCTGCGCGCTGCTCGCGCATCTGAG 21366
Qy 541 GGCACCGGAGTGGCGCGCGCGCGCGAGTCCGCTGATGCTATGCTGCGACTGCGCCCATG 600
Db 21367 TCGAGGAGGCTCTGCGCGCTCAAGCTGATGATGATGCTGCGCTTCCATCGCGCGAGT 21426
Qy 601 CAGCCCTCGGAGGAGTGTCTGAGCTGTGAGGCGGAGCATCAAGCGCGAGCGCGTCCG 660
Db 21427 GAGTCGATTCGCGAGCGAGCTCTCGATCTCTGTGCTGAGCGCGCGCTGAGCGG 21486
Qy 661 GTGCGCTTCTTCTCCAGCTGTGAGGCGGACCTGTGAGCAACCAACCTGTGAGCGCG 720
Db 21487 GTCCGCTTCTTCTCCAGCTGTGAGGCGGCGCGCGCATGAGCGGAGCGAGCTCGACG 21546
Qy 721 TACTGTATCCGCAACCTGACAGCGCGGTCCGTTTCAAGGATGCCCTCCAGGCGCGTGG 780
Db 21547 TACTGTATCCGCAACCTGACAGCGCGGTCCGTTTCAAGGATGCCCTTGTGCAAGCG 21606
Qy 781 GATGACGACACCGCGCTCTTGTGAGTCAAGCGCGCGACCCCAACCTCTGCTGCGCGCAT 840
Db 21607 GCGGAGGACATCGCTTCTTGTGAGGAGTGAAGCGCGCGCATGCTGTGAGCTTGGCT 21666
Qy 841 GAAAGACACCGGAGAGACCGCGCGGAGAGCTGACCGCATGCGGAGCTCGCGCGCG 900
Db 21667 CACGAGCTCTCGAAGCGGTGAGCGCTGCGGCGCGGTGTGCTGTCTGTGAGCGAG 21726
Qy 901 GACAAGACACCGCGCGCTTC 921
Db 21727 GAAAGGATCTTACGCGCTTC 21747

RESULT 15
US-08-458-076A-6
Sequence 6, Application US/08458076A
Patent No. 5658425
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Ukner, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
antipathogenic substances
NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:
ADDRESS: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,076A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIORITY DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Rimer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076A-6

Query Match 37.5%; Score 345; DB 1; Length 28958;
Best Local Similarity 61.7%; Pred. No. 9,9e-52;
Matches 568; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

Qy 1 GTCTTGATGTTTCCGAGAGGAGTGGAGTGGCGGAGTGGCTGTGCTGCGC 60
Db 20830 GTCTTGATGTTTCCGAGAGGAGTGGAGTGGCGGAGTGGCTGTGCTGCTCAT 20889
Qy 61 TCTCTCGGATGTTGCGCGCGGATGCAAGCGCTGCGAGAGGCTTGGCGCGTGG 120
Db 20830 ACTCGCGGATGTTTCCGAGAGGATGCAAGCGCTGCGAGCGCGCTGCGCGCACG 20949
Qy 121 GATGATGTCAGCCTGTGCTGTTTCAAGCTCATGATGTTTGTGCTGCTGTGAGCTTCC 240
Db 20950 GACTGTGTCAGCCTGTGCTGTTTCAAGCTCATGATGTTTGTGCTGCTGTGAGCTTCC 21009
Qy 181 GATGATGTCAGCCTGTGCTGTTTCAAGCTCATGATGTTTGTGCTGCTGTGAGCTTCC 240
Db 21010 GACGTGTCAGCGCGGCTTCTGATGATGATGCTCGCTGCGCCCGCTGTGAGCTCC 21069
Qy 241 TACGATATGAACCCGACCGCGCTCTTGGCCATTCCGAGGCGAGATCGCGCGCGCAT 300
Db 21070 ATGGGCGGTGAGCGCCGACCGGCGGTGCGCATAGCGAGGCGAGATCGCGCGCGCTGT 21129
Qy 301 GTGTGTGGGGCGGTGAGCGCTGAAAGACGCGGCGAAGACTTGTGCGCTGCGGACCGCGG 360
Db 21130 GTGGCGGGCGCGGTGTGCTGCGTGAAGCGCTGCGCAACCTGTGCGCTGTGCGACCGT 21189
Qy 361 CTGGCGCTGTGCGGCGCGCGCGCGCATGAGCTCAATGCGCGCTGCTGCGCCGAGAGTG 420
Db 21190 CTGTGTGAGCTCGCCCGCGGCGGCGGCGCATGCGCGGTGAGCTGCGGAGCGCGAGT 21249
Qy 421 GAGCAGCTCATTTGTGAGCGGTGAGGCGGCGGCGGTGTGAGGCGCGGCTCAACGCG 480

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-342-6

Query Match 37.5%; Score 345; DB 1; Length 28958;
Best Local Similarity 61.7%; Pred. No. 9,9e-52;
Matches 568; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

1 GTCCTGCTGTTTCCCGACAGAGGTGAGTGGAGCGGAGTGGGTGCTGCTGCTGCC 60
20830 GTCCTGCTGTTTCTTCTGCGGACAGCTGCAATGGAGATGGCCCTCTCCCTGCTCAT 20889
61 TCCTCTCCGAGTGTTCGCCGCGGATGACAGCGTGGAGAGGCTTGCGCGCCGAGTGG 120
20890 ACCTCGCCGGTCTTCCGGGACAGCTGAAAGCTGGAGCGCCCTCGCGGCCACAGTG 20949
121 GATTGCTGTGTGTGACATCTTCCGCCGAGACCGGGGAGTGGGTGGAGCGCGGCC 180
20950 GACTGTGCTGTGCTGCTGCGGCTGCTCCGAGGAGAGGCGCGCCCGCTGCACTGATC 21009
181 GATGTGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
21010 GACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21069
241 TACGATATGACACCGACCGCGGTCTTGGCATTCCAGAGGCGAGATCGCGCGCGCAT 300
21070 ATGGGCGTGCAGGCGGACCGCGGTGCTGCGGCTATAGCCAGGCGAGATCGCGCGCTG 21129
301 GGTGTGTGGGGGCTGAGACCTGAAAGACCGCGGCAAGACTGTTCCGCTGGAGCGCGGG 360
21130 GTGGCGGGGCGGTGCTGCTGCTGAGAGGCTGCAAGCTGTGGCGCTGCGGACCGGTGG 21189
361 CTGGCGCTGTGTGGGGGCGGCGGCGGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
21190 CTGCTGAGAGCTCCCGGCGGAGGCGGCGGCTGCGGCTGAGTGGCGGAGCGCGAGTTC 21249
421 GACCACTCATTTGTGAGCGGTGCGGCGGCGGTGTGGGTGGCGCGGCTGCAACGCGCCC 480
21250 GCACGCGGCTCTTCCAGCGCTATGCGGATCGGCTCTCATCGGCGGATCAACACGCTC 21306
481 CGCTCAACCGCGCTCTCGGCGGATGCGGCGGCTGAGCGAGTGTGCTGCTGCTGCTGCT 540
21307 CGTTTCAACGATCTTCCGCGGAGCGGCTTGGCGCGGCTGCTGCTGCGGATCTGGAG 21366
541 GGCACCGGGGTGGGCGGCGGCGGATGCGGCTGAGTATGCTGCTGCTGCTGCTGCTGCTG 600
21367 TCCGAGGGGCTCTTCCGCTCAAGCTGAGTATGAGTATGCTGCTGCTGCTGCTGCTGCT 21426
601 CAGCCCTGTGGGAGAGTGTGCTGAGCTGTGAGGAGATCAAGCCCGAGCGCTGCGGCG 660
21427 GAGTCAATTTGCGACGAGCTCTGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 21486
661 GTCGCGCTTCTTCTTCCAGCGGTGAGGAGGACCTGCTGAGCAACACCTTGGAGCGCGCC 720
21487 GTCGCGCTTCTTCTTCCAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 21546
721 TACTGTACCGCAACTGACACAGCGGCTGCTTTCAGCGATGCGGCTGCGAGCGCTGGAG 780
21547 TACTGTACCGCAACTGCGGAGCGGCTGCTTCTGAGAGTGGAGCGCCAGTCTGTGCTG 21606
781 GATGAGGAGACCGCGCTTCTTCTGCTGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGG 840
21607 GCGGAGAGATGCTGCTTCTTCTGAGAGTGGAGCGCCAGTCTGTGCTGAGCTTGGGCT 21666
841 GAGAGACACCGAGAGACACCGCGGAGAGAGTATCGGCGGATGCGGCGGCGGCGGCGG 900
21667 CAGAGGCTCTTCAAGAGGCTGAGAGCGCTCGGCGGCGGCTGCTGCTGCTGCTGCTGCT 21726
901 GACAGGACACCGCGGCTTTC 921

Db 21727 GAGAGGATCTACGCGGCTTC 21747

RESULT 14

US-08-457-646A-6
Sequence 6, Application US/08457646A
Patent No. 5679560

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ring
APPLICANT: Hammer, Phillip E.
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457, 646A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457, 205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08-Jun-1994

ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689

INFORMATION FOR SEO ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-646A-6

Query Match 37.5%; Score 345; DB 1; Length 28958;
Best Local Similarity 61.7%; Pred. No. 9,9e-52;
Matches 568; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

1 GTCCTGCTGTTTCCCGACAGAGGTGAGTGGAGCGGAGTGGGTGCTGCTGCTGCC 60
20830 GTCCTGCTGTTTCTTCTGCGGACAGCTGCAATGGAGATGGCCCTCTCCCTGCTCAT 20889
61 TCCTCTCCGAGTGTTCGCCGCGGATGACAGCGTGGAGAGGCTTGCGCGCCGAGTGG 120
20890 ACCTCGCCGGTCTTCCGGGACAGCTGAAAGCTGGAGCGCCCTCGCGGCCACAGTG 20949
121 GACTGTGCTGTGTGACATCTTCCGCCGAGACCGCGGAGATGGGTGGAGCGCGGCC 180

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip B.
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF INVENTIONS: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6

Query Match 37.5%; Score 345; DB 1; Length 28958;
Best Local Similarity 61.7%; Pred. No. 9.9e-52;
Matches 568; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

QY 1 GTCTGTGTTTCCCGACAGGAGGAGGAGTGGGCGGAGATGGATGGTGTGCTGTGCTGCC 60
DB 20830 GTCTGTGTTTCCCGACAGGAGGAGTGGGCGGAGATGGATGGTGTGCTGTGCTGCC 20889
QY 61 TCCTCTCCGGTGTTCGCCCGGCGAGTGCAGCGCTGCAGAGAGCTGTGCGCGCGGAGTGC 120
DB 20890 ACCTCGCGGCTCTTCGCGGACAGCTGAGCGTGCAGCGCGCGCGCTGTGCGCGCGG 20949
QY 121 GACTGTGTGTGTGACATCTTCGCGCGGAGCGCGGAGATGGCGTGTGGAGCGCGGCC 180
DB 20950 GACTGTGTGTGTGACATCTTCGCGCGGAGCGCGGAGATGGCGTGTGGAGCGCGGCC 21009
QY 181 GATGTGGTCCAGCTGTGCTGTTCAGGCGTATGTCTTTGAGTGTGTGTGAGCGCTTC 240
DB 21010 GACTGTGGTCCAGCTGTGCTGTTCAGGCGTATGTCTTTGAGTGTGTGTGAGCGCTTC 21069
QY 241 TACGGTATCGAACCCGAGCGGCTCTTGAGCATTTCCAGGAGCGAGATCGCGCGCGCAT 300
DB 21070 ATGGGCGTCCAGCGCGGAGCGCGGTGTGCGGCATATGCGAGGCGAGATCGCGCGCGCTGT 21129

QY 301 GTGTGTGGGCGCTGAGCTGAAGGACGCGGAGAGACTGTGCGCTGCGAGCGCGGCG 360
DB 21130 GTGTGTGGGCGCTGAGCTGAAGGACGCGGAGAGACTGTGCGCTGCGAGCGCGGCG 21189
QY 361 CTGGCGGTGTGTCGCGGCGCGGCGGAGTGGCTTCAGTCCGCTTCCTGCGAGAGTGC 420
DB 21190 CTGGCGGTGTGTCGCGGCGCGGCGGAGTGGCTTCAGTCCGCTTCCTGCGAGAGTGC 21249
QY 421 GAGCAGCTCATTTGTGAGCGGTGTGCGCGGCGGCTGTGTGGGTGTGGCGGCTCAACGCGCC 480
DB 21250 GCACGCGCGCT---CCAGCGCTATGCGCATGCGCTTCATCGCGGCGCATCAACGCGCT 21306
QY 481 CGCTCACCGCGCTGTGCGGCGGAGTGCAGAGCGGTGAGCAAGATGCGTGAAGTGTGCC 540
DB 21307 CGTTTCAGACGATCTTCGCGGAGAGCGCGCGCTTCGCGCGCGCTTCCTGCGAGTGTGAG 21366
QY 541 GGCACCGGAGTGTGCGGCGCGGCGGAGTCCGCGTGAATGCTTCGCACTGCGCCCATGTG 600
DB 21367 TCCGAGGGGCTTTCGCGCGCTTCAGTGAATGAGTTCGCTTCGCACTGCGCCCATGTG 21426
QY 601 CAGCCCTGCGGAGAGATGTGTCGAGCTGTGCGGAGATGACCCGCGAGCGCTTCGCGC 660
DB 21427 GAGTGTATTCGCGAGAGCTCTTCGATCTCTGTGTGAGCTTCGAGCGCGCTTCGAGCGG 21486
QY 661 GTGCGGCTTCTTCAGCGGTGAGGCGCACTGCTGAGCACCAACCTTGAAGCGCGCG 720
DB 21487 GTCCGCTTCTTCAGCGGTGAGGCGCACTGCTGAGCACCAACCTTGAAGCGCGCG 21546
QY 721 TACTGTATCCGCACTTCGAGCGCGGCTTCCTTCAGGATGCGGCTTCGAGCGCGCTGAGG 780
DB 21547 TACTGTATCCGCACTTCGAGCGCGGCTTCCTTCAGGATGCGGCTTCGAGCGCGCTGAGG 21606
QY 781 GATGACGAGACCGGCTTCTTCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 840
DB 21607 GCGGAGAACATCGCTTCTTCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 21666
QY 841 GAAACACCGAGAGAGACCGCGGAGAGCGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 900
DB 21667 CAGGAGCTCTTCGAGAGCGTTCGAGAGCGGCTTCGAGAGCGGCTTCGAGAGCGGCTTC 21726
QY 901 GACAACGACCGCGCGCTTC 921
DB 21727 GAAAGGAGTCTACGCGGCTTC 21747

RESULT 12
US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip B.
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF INVENTIONS: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

QY 418 GTGAGACGCTCATTTGTTGAGCGGTGGCGGGCGGTTGTGGTGGCGGCTCAACGGC 477
 DB 8682 GAGAACTGCTCGCGCGACGACAGCTGGACCGGCGGCTGGAGGTCCCGCGTCAACGGC 8741
 QY 478 CCCCGCTTCACCGCGCTCTCGGGGGATGCCAGGCGGTGACGAGGTGCTGGCTACTGT 537
 DB 8742 CCCGACCGCTGTGGTGGCGGAGCGCCAGCGCGCGGCGGCTTCTGGAGTACTGC 8801
 QY 538 GCCGCGACCGGGGTGCGGGCGCGGCGGATCCCGGTGACTATGCTTGCATCGCCCAT 597
 DB 8802 GAGGCGCTGGGCACTCCCGCGCGCGGCGGATCCCGGTGACTATGCTTGCATCGCCCAT 8861
 QY 598 GTGACAGCCCTCGCGGAGAGGTGCTGAGCTGTGAGGAGACATCAGCCCGCAGCGCTCC 657
 DB 8862 GTGAGAGCCGCTGCGCGACGAACTGTGTCAGGCGCTGCGCGGAGTACACCCGCGACGCGGC 8921
 QY 658 GGGGTGCGCTTCTTCTTCACAGGTGAGAGGCACTGGCTGAGACACACACCTTGAACGCC 717
 DB 8922 GAGGTGCGCTTCTTCTTCACAGGTGAGAGGCACTGGCTGAGACACACACCTTGAACGCC 8981
 QY 718 GCGTACTGTACCGCAACTGACACAGCGGTCGCTTTCAGCGATGCGCTTCAGCGCGCTG 777
 DB 8982 GCGTACTGTACCGCAACTGACACAGCGGTCGCTTTCAGCGATGCGCTTTCAGCGCGCTG 9041
 QY 778 GCGGATGACGAGACCGCGTCTTCTTCAGAGTACAGCCCGCACCCCTGCTCCCGCGC 837
 DB 9042 ACCGACGAGGATACGAGGAGCTTTCATGAGGTGAGGTCGACCGCGGCTGGAGCGCTG 9101
 QY 838 ATGGAAGACACACCGGAGACACCGCGGAGAGTACACCGCGGATGCGGAGCTTCCGCGC 897
 DB 9102 GTCCAGAGACCTCTCGACGACGCGGAGTCCGAGCGCGGCTGCTGGAGCGCTGGAACGC 9161
 QY 898 GCGGACACGACGACCGCGCGCTTC 921
 DB 9162 GACGCGGCGGACCGCGGCTTC 9185

RESULT 8

US-08-439-009A-1

Sequence 1, Application US/08439009A

Patent No. 6004787

GENERAL INFORMATION:

APPLICANT: Donadio, S

APPLICANT: Mcalpine, J B

TITLE OF INVENTION: Method of Directing Biosynthesis of

TITLE OF INVENTION: Specific Polypeptides

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSER: Steven P. Weinstein

STREET: Abbott Laboratories D377/APed-2 One Abbott

CITY: Abbott Park

STATE: IL

COUNTRY: US

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/439,009A

FILING DATE: 11-MAY-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Casuto, Dianne

REGISTRATION NUMBER: 40,943

REFERENCE/DOCKET NUMBER: 4952-US-D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847-938-3137

INFORMATION FOR SEQ ID NO: 1:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 LENGTH: 11219 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Saccharopolyspora erythraea
 STRAIN: NRRL 2338
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 744..6659
 OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
 OTHER INFORMATION: MODULE 1"
 OTHER INFORMATION: /label= FUNCTION
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 744..11219
 OTHER INFORMATION: /function= "gene= "eryA"
 OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
 OTHER INFORMATION: 6-deoxyerythromolide B"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 744..1868
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: acyltransferase domain 1 of module 1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1998..2198
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: acyl carrier domain 1 of module 1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 2250..3626
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 3831..4811
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: acyltransferase 2 domain of module 1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 5574..6125
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: beta-ketoreductase domain of module 1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 6369..6626
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: acyl carrier domain 2 of module 1"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 6678..11219
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: module 2"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 6678..8066
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 8262..9305
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: acyltransferase domain of module 2"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 9906..10454
 OTHER INFORMATION: /function= "approximate span of
 OTHER INFORMATION: beta-ketoreductase domain of module 2"

```

US-07-642-734C-1
Sequence 1, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckerts, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..6659
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B""
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..1868
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1".
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"

```

Db 550 GTGAGGGGTGCAACCGCCCGATCCCGGCTGACACAGGCGGCACTGCGCGAG 609
 Qy 598 GTGACAGCCCTGCGGAGAGATTGCTGAGCTGCTGAGGAGACATACCGCCGAGCGTCC 657
 Db 610 GTGAGAGCGGTTGGGGGTGATCTGCTGAGAGTGTGCGCCCGGTCGCCGCCAGCGGCC 669
 Qy 658 GGGGTGCGCTTCTTCTTCACGCTGAGAGGGACCTGGCTGACACACACCTTGAAGCC 717
 Db 670 GACATCCCGTTTACTGACGCTGACGCGGCGGCTGCTGACGCGACCGAGCTGAGCGCG 729
 Qy 718 GCTACTGTGATCGCAACCTGACACAGCGGCTCGTTTACGAGATGCGGCTCGAGGCGCTG 777
 Db 730 ACCTACTGTGATCGCAACCTGACACAGCGGCTCGTTTACGAGAGTTCAGCGGCGACACAGGCGCTG 789
 Qy 778 GCGGATGACGAGACCGCGCTCTTCTGCTGAGTCAAGTCAAGCCCGACCCCTGCTCGCGCC 837
 Db 790 ATGCGCGAGCGGAGACGAGTCTTCTGAGAGAGCGCGCATCCATGCTGGCGGTGGCG 849
 Qy 838 ATCGAAGACACACCGGAGACACCGCGGAAAGCTCACCGCGATCGGAGCGCTCCGCGCC 897
 Db 850 CTGAGACGAGCGGTCAACGACGCGCGGACCGAGCGGCGGTGCTCGGAGACCTTGGCGCGC 909
 Qy 898 GGGGACAGACACCGCGC 915
 Db 910 CGCCAGCGGCTCTCGC 927

RESULT 6

US-08-997-467-29

Sequence 29, Application US/08997467

Patent No. 6200813

GENERAL INFORMATION:

APPLICANT: Katz, Leonard

APPLICANT: Staszi, Diane L.

APPLICANT: Summers Jr., Richard G.

APPLICANT: Ruan, Xiaolan

APPLICANT: Pereda-Lopez, Ana

APPLICANT: Kakavas, Stephan J.

TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES

TITLE OF INVENTION: AND RECOMBINANT METHODS FOR MAKING SAME

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Rd.

CITY: Abbott Park

STATE: Illinois

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: PasteSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,467

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/858,003

FILING DATE: 16-MAY-1997

ATTORNEY/AGENT INFORMATION:

NAME: Diane Casato

REGISTRATION NUMBER: P-40,943

REFERENCE/DOCKET NUMBER: 4952.US.P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (847)-938-3137

TELEFAX: (847)-938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 1010 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
 US-08-997-467-29

Query Match 39.4%; Score 362.8; DB 4; Length 1010;
 Best Local Similarity 63.7%; Pred. No. 9.3e-55;

Matches 585; Conservative 0; Mismatches 327; Indels 6; Gaps 2;

Qy 1 GTCTTGCTGTTTCCCGAGAGGATGAGAGTGGAGCCGAGATGAGTGTGCTGTGCTGCGC 60
 Db 13 GTGTTGTTGTTCCCGGAGGAGGCTGCGAGTGGGCGGAGATGGCCAGAGGGGCTGTGGAG 72
 Qy 61 TCTCTCCGCTGTTGCGCGGATGACAGGGGTGAGAGAGCTTGGCGCGGTGGAG 120
 Db 73 CGGTCCGCGCGCTTCCGAGATGCGCGGAGCTGATGAGACCGCGCTGCGCGCTGAC 132
 Qy 121 GACTGTCTGTGATGACATCTGAGCGCGGAGACGCGGGGAGATGCGGTGTGGAGCGGCGC 180
 Db 133 GGTGTGCTGTGCTGAGCGCTGCTGCGGGAGAACGAGCGGCTTCTGCTGACCGGATC 192
 Qy 181 GATGTGTCAGACCTGTGCTGTTCAAGCTGATGTGTCTTTGAGCTGCTGTGCGCTTCC 240
 Db 193 GAGTGTGTCAGCGGATGCTGTTCAAGATGATGTCTGCTGCGCGGCTGTGGCGTGG 252
 Qy 241 TACGTTATGAACCGACCGCGTCTTGGCACTTCCAGGGGAGATGCGCGCGGCAAT 300
 Db 253 CTGGGGGTGAAACCGCGCGCGGTCTGCGGCACTCGACAGGTGAGATCCCGCTGCCAT 312
 Qy 301 GTGTGTGGGGGCTGAGCCTGAAGAGACGCGGGGAGACGTTGCGCTGTGGACCGGGCG 360
 Db 313 GTGCGCGGTGCTGTGCTGAGACATGCGCGCGGATGATGCGCTTGGAGTGGGG 372
 Qy 361 CTGGCGCTGTGTCGGGGCGGGCGGCGCATGAGCTGCGCTGCTGCCAGAGAGTGG 420
 Db 373 TGGCTGAGACTGGCGGGGAGAGGGCGGCAATGTGGGTGCTCCATGCTGGCGAGAGAGTGG 432
 Qy 421 GAGCACTCATTTGTGAGCGGTGGCGGGCGGCTGTGTGAGTGGCGGCTCAACGCGCCC 480
 Db 433 CGGCGCGGCT---GGTGAAGTGGGGGAGCCGTGCGCGCTGCGCGCGCTCAACGCGCCC 489
 Qy 481 CGCTCACCGCGCTGCTCGGGGAGATGCCAGGGGGTGAAGAGTGTGCGCTGACTGTGCGC 540
 Db 490 GGTTCCTGCGCGGTGCGAGCGACCCGAGGGGCTGGCCGAATCGGTGGCGCTGTGACC 549
 Qy 541 GGCACCGGGGTGCGGGCGCGCGAGATCC---GATCACTATGCTCGACTGCCCCCAT 597
 Db 550 GGTGAGGGGGTGCAGCGCGCGCGCATCCCGGCGTGAACGCGGGGCGACTGCGCGAG 609
 Qy 598 GTGACAGCCCTGCGGGAGAGTGTGCTGAGCTGTGGGGGACATGACCCGAGCGCTCC 657
 Db 610 GTGAGCGCGTTCGGGCTCATCTGCTGAGGTGTGGCCCGGCTGCGCCCGAGCGGCGC 669
 Qy 658 GGGGTGCGGTTCTTCTCCACGCTGAGAGGCACTGTGGTGAACACACACCTTGAAGCGCC 717
 Db 670 GACATCCCGTTTACTGACGCTGACCGCGGGCTGTGAGGACCGAGCTGAGCGCG 729
 Qy 718 GCTACTGTGATCGCAACCTGACACAGCGGCTCGTTTACGAGATGCGGCTCGAGGCGCTG 777
 Db 730 ACCTACTGTGATCGCAACCTGACACAGCGGCTCGTTTACGAGAGTTCAGCGGCGACACAGGCGCTG 789
 Qy 778 GCGGATGACGAGACACCGCGCTTCTTGTGAGTGAAGTGAAGCTGACCCGACCCGCTGTCCCGCGC 837
 Db 790 ATGCGCGAGGAGCAAGAGTCTTCTTGAAGAGAGCGCGCATCCATGCTGAGCGGTGG 849
 Qy 838 ATCGAAGACACACCGGAGACACCGCGGAAAGCTCACCGGATGAGGAGCTTGGCGCGCGCC 897
 Db 850 CTGAGACGAGCGGTACCGACCGCGGACCGAGCGGGCGGTCTGGAGACCTTGGCGCGC 909
 Qy 898 GGGGACAGACACCGCGC 915
 Db 910 CGCCAGCGGCTCTCGC 927

RESULT 7

NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15165..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

Query Match 41.84; Score 385.4; DB 3; Length 20235;
Best Local Similarity 65.64; Pred. No. 1.1e-58;
Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;
QY 1 GTCTTCGTCTTCCGACAGAGGTGGAGTGGCCGCGATGGTGTGCTGTCTGCTGCC 60

Db 1696 GTCTTCGTCTTCCGACAGAGGTGGAGTGGCCGCGATGGTGTGCTGTCTGCTGCC 1755
QY 61 TCCTCTCCGGGTGTTGCCCGCGAGTGCAGAGGCTCTGGCCCGGTGGG 120
Db 1756 TCGTCCCGGTGTTGCCCGGGAAGTCCGGGCTGCGAGATCATGGCCCCGATGCG 1815
QY 121 GACTGATCTGATGGATCATCTGCGCGGAGCGCGGGGAGATGGGTGAGGAGCGGCC 180
Db 1816 GACTGAAAGGTCTCGAGAGTCTGCTCAGAGCGCGGGGGCGCCCGGCTGAGACGGGTG 1875
QY 181 GATGTGATCCAGCTGTGCTGTTCAACGTCATGTGTCTTTGGCTGTCTGTGTGCTTC 240
Db 1876 GACGTGTGACAGCGGTGTGTTGCGGGTGAATGTGTGCTGTGCGAGCTGTGGCGCTG 1935
QY 241 TAGCGTATCGAACCCGAGCGGTCTTTGGCCATTCCAGGAGGAGATGCGGCGCAT 300
Db 1936 TAGCGCTGTGAGACCCGCGCGGTCTGTGGGCGCATGCAAGGCGAGATCGCCCGCGAC 1995
QY 301 GTGTGTGGGAGCGTGAAGCTGAAGACGCGGGAAGACTGTGGCGTGCAGCGCGGCG 360
Db 1996 GTGCGCGGAGCGCTCACTTGAAGACGCGGAGAGCTGTGTGGCGCGACGCGCTG 2055
QY 361 CTGGCCGCTGTGCGGGGCGCGGCGCATGTGCTCATGTGCGCTGCTGCCAGAGGTG 420
Db 2056 ATGCGGTGCTCTCGCGGAGAGGCGGCGCATGTGCGCTGCGCTGAGGCGCGCGGTG 2115
QY 421 GAGCAGCTCATTTGTGAGCGGTGTGGCGGGCGGCTGTGTGGGTGGCGCGGTCAAAGCCCC 480
Db 2116 CCGGAGCGCTGCGG---GCCGTGCGAGAGACCGGCTCTGTGTGGCGCGGTCAAAGCTCC 2172
QY 481 CGCTCACCGCGGTCTCGGGGGGATGCGAGGCGGTGAGCGAGGTCTGGCGTACTGTGTC 540
Db 2173 CGTGTGTGTGTCTTCGCGAGAGCGCGCGGCTGCTGCGGCGCTTTCCAGAGACTGTGCG 2232
QY 541 GGCACCGGGGTGCGGGCGCGCGGATGCCGCTGACTATGCTCTGCACTGCCCGCATGTG 600
Db 2233 GCGGAGGGGATCGGGTCCGCGCATGCGAGCTGAGCTACGCTCGACATCGCGCGAGATC 2292
QY 601 CAGCCCTGCGGAGAGATGTGTGAGAGCTGTGCGGGAGATAGCCCGGAGCGGTTCGGCG 660
Db 2293 GAGCGGGTCCGCGAGAACTCTCGAAGACGCGCGAGATGCGCGCGCGCGCGCG 2352
QY 661 GTGCGGTCTTCTCACGCGTGAAGGCACTGTGTCAGACACACCACTGAGCGCGCG 720
Db 2353 GTGAGTTCACCTCACTGTGAGTGTGCGGCTTATGAGCGGACCGAGCTGATGCTCCG 2412
QY 721 TACTGTACCGCACTGTGACAGCGGTCTGTTCAAGAGATGCCGTCCAGGCGCTGGCG 780
Db 2413 TACTGTACCGCACTGTGCGGAGAGCGAGCGGTCTGCGGAGCGCGGTGACGCGGTGGCG 2472
QY 781 GATGACGAGACCGGTCTTCTGTGAGAGTCAAGCCCCACCCACCTGTCTCCCGCATC 840
Db 2473 GAGTGGGATGAGAGCGGTCTTCTGAGAGTCAAGCCCCACCCACCTGTCTCCCGCATC 2532
QY 841 GAAAGACACGAGAGACACGCGGAGAGCGTCAACCGGAGTGG 883
Db 2533 GAGAGGCGGTGAAAGAGGCTGACGCTCCGAAGCGCGGTG 2575

RESULT 3
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Bellach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428, 517
; CURRENT FILING DATE: 1999-10-28

```

QY 301 GTGTGTGGGGCGCTGAGCCTGAAGAGCGCGGCAAGACTGTTCGCTGCGCAGCCGCGCG 360
DB 1996 GTGCGCGGGGCGCTCAAGTTGAGAGACGCGCGCAAGCTGCTGTGGGGCGCGACCGCGCTG 2055
QY 361 CTGGCCGCTGTGGGGGGCGCGCGGATGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 2056 ATGCGGTGCTCTCCGGGAGGGCGCGATGCGCGCTGCGCTGCGCGCGCGCGCGCTG 2115
QY 421 GAGCAGCTCATTTGTGAGCGGTGAGCGGGCGGTGTTGGTGGGCGGTCAACGGCGCC 480
DB 2116 CCGGAGCGCTGGG---GCCGTGGCAGAGACCGGCTCTCGGTGGCGCGCGGTCAACGGTCC 2172
QY 481 CGCTCAACCGCCGCTTCTGCGGGGATGCCGAGCGGTGAGCAGAGTGTCTGCTGCTGCTG 540
DB 2173 CGGTGCGTGTGCTCTCCGGGAGCGCGCGCGCGCTGCGCGCGGTTCCTCGAGGACTGCGCG 2232
QY 541 GGGACCGGGGTGGGGCGCGCGCGGATCCCGGTGACTATGCTGCTGCACTGCCCGCATG 600
DB 2233 GCGGAGGGGATCCGCGCTCCGCAATCGACGTGAGACTACGCTCGCACTGCGCGCAGATC 2292
QY 601 CAGCCCTGCGGGAGAGATTGTGAGCTGCTGAGGGGACATCAGCCCGCAGCGCGTCCG 660
DB 2293 GAGCGGGTCCGCGAGAACTCTCGAAACGACCGCGGACATCCGCGCGCGCGCGCGCG 2352
QY 661 GTGCGCTTCTTCTCAACGCTGAGAGGCACTGCTGAGACACCAACCTTGAACGCGCG 720
DB 2353 GTGACGTTCACTCCACTGCTGAGTGGCGGTCTATGAGCGGACCGAGCTGATGCGCG 2412
QY 721 TACTGTATCCGCACTGACACGCGGCTGCTTCAAGCATGCTGCTGCGCGCGCTG 780
DB 2413 TACTGTATCCGCACTGACACGCGGCTGCTTCAAGCATGCTGCTGCGCGCGCTG 2472
QY 781 GATGACGACACCGCGCTTCTGCGAAGTCAAGTCAAGCGCCCGACCGCTGCTGCGCGCAT 840
DB 2473 GAGTGGGATACGACCGCTTCACTGAGGTACGCGCGCATCCGCTGCTGCTGCAAGCGCGT 2532
QY 841 GAAGACCAACCGAGACACCGCGCGGAAAGTCAACCGCGCATG 883
DB 2533 GAGGAGCGGTGAGAGAGGCTGACGCGTGCAGGAAGCGCGGTG 2575

RESULT 2
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AB6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943

```

```

; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19
; OTHER INFORMATION: /function= "gene eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 19..4470
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 97..1482
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1693..2670
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 3406..3921
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4171..4428
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4471..10722
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: module 4"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 4471..5847
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 3"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 6054..7026
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 4"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 7165..9216
; OTHER INFORMATION: /function= "approximate span of
; OTHER INFORMATION: dehydratase and enoylreductase domains m"
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 9433..9984
; OTHER INFORMATION: /function= "approximate span
; OTHER INFORMATION: beta-ketoreductase of module 4"
; FEATURE:

```



```

OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9453..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatlo"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature

```

```

LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19145..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

Query Match      41.8%; Score 385.4; DB 1; Length 20235;
Best Local Similarity 65.6%; Pred. No. 1.1e-58;
Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;

Qy      1  GCTTCGTGTTCCCGACAGGAGTGAGGCGCGGAGTGGTGTGCTGCTGCC 60
Db      1696  GTCTTCCTGTTCCCGGAGGAGGTTGCAATGGGCGCGATGGTGCAGAGCTGCTG 1755
Qy      61  TCTCTCCGAGTGTGCGCCCGGAGATGAGGCGTGGAGAGCTTGGCGCCGCGGAG 120
Db      1756  TCGTCCCGAGTGTGCGCCCGGAGATCCCGGCGCTGGAGAGTGCATGCGCCGATGCA 1815
Qy      121  GACTGCTGTGTGTGATCATCTGCGCGGAGACCGGAGGAGTGTGGAGCGGGCC 180
Db      1816  GACTGAGAGTCTCCACAGTGTGCTGCTGAGGCGCGGAGCGCGGCTGAGACGAGT 1875
Qy      181  GATGTGTCAGAGCTGTGCTGTTCAGCGTCAATGATCTTTTGGCTGCTGTGGCGTTCC 240
Db      1876  GACGTGTGACAGCGGAGTGTGTTGCGCGATGATGTGTCTGCGGAGCGTGTGGCGCTG 1935
Qy      241  TACGGTATGAAACCCGACCGCGGTCTTGGCCATTTCCAGGAGGAGATCGCGGCGGCGAT 300
Db      1936  TACGGGTGAGACCCCGGCGGTGTGTGGGCACTGCGACGAGGAGATCGCGCGGCGAC 1995

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 39.8967 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068

Perfect score: 921 gcttcgcttccgcgcgaca.....acaacgacacccgcgcttc 921

Sequence: 1 gcttcgcttccgcgcgaca.....acaacgacacccgcgcttc 921

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/prodata/1/ina/PCOTUS_COMB.seq:*

6: /cgn2_6/prodata/1/ina/backfile1.seq:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385.4	41.8	20235	1	US-07-642-734C-3
2	385.4	41.8	20235	3	US-08-439-009A-3
3	373.8	40.6	50937	4	US-09-428-517-1
4	362.8	39.4	1010	3	US-08-858-003-29
5	362.8	39.4	1010	3	US-09-078-166-29
6	362.8	39.4	1010	4	US-08-997-467-29
7	349.6	38.0	11219	1	US-07-642-734C-1
8	349.6	38.0	11219	3	US-08-439-009A-1
9	348	37.8	80161	3	US-09-370-987A-1
10	348	37.8	80161	4	US-08-258-261B-6
11	345	37.5	28958	1	US-08-456-837-6
12	345	37.5	28958	1	US-08-457-342-6
13	345	37.5	28958	1	US-08-457-342-6
14	345	37.5	28958	1	US-08-457-646A-6
15	345	37.5	28958	1	US-08-458-076A-6
16	345	37.5	28958	1	US-08-764-233A-4
17	345	37.5	28958	1	US-08-457-335A-6
18	345	37.5	28958	1	US-08-729-214-6
19	345	37.5	28958	3	US-09-028-934-6
20	345	37.5	49377	1	US-08-764-233A-1
21	344.4	37.4	44377	2	US-08-804-227C-7
22	344.4	37.4	44377	2	US-08-804-198-1
23	342.2	37.2	43280	2	US-08-804-227C-1
24	319	34.6	441529	4	US-09-103-840A-1
25	318	34.5	13842	4	US-09-105-537-30
26	318	34.5	13842	4	US-09-105-537-5
27	318	34.5	38506	3	US-09-320-878-19

28	311	33.8	11220	4	US-09-105-537-32	Sequence 32, Appl
29	308.6	33.5	1035	3	US-08-858-003-30	Sequence 30, Appl
30	308.6	33.5	1035	3	US-09-078-166-30	Sequence 30, Appl
31	308.6	33.5	1035	3	US-08-997-467-30	Sequence 30, Appl
32	308.6	33.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
33	306.8	33.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl
34	306.8	33.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
35	305.8	33.2	4689	4	US-09-105-537-34	Sequence 34, Appl
36	302	32.8	1881	4	US-09-434-288-5	Sequence 5, Appl
37	298.4	32.4	68750	3	US-09-335-409-1	Sequence 1, Appl
38	298.4	32.4	68750	4	US-09-568-102-1	Sequence 1, Appl
39	298.4	32.4	68750	4	US-09-567-969-1	Sequence 1, Appl
40	298.4	32.4	68750	4	US-09-568-480-1	Sequence 1, Appl
41	298.4	32.4	68750	4	US-09-568-486-1	Sequence 1, Appl
42	298.4	32.4	68750	4	US-09-568-472-1	Sequence 1, Appl
43	298.4	32.4	68750	4	US-09-567-899-1	Sequence 1, Appl
44	296.8	32.2	71989	4	US-09-443-501A-2	Sequence 2, Appl
45	293.8	31.9	13987	2	US-08-804-227C-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-07-642-734C-3
Sequence 3, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckert, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722

QY 721 GGGAGGGGCTGGGATGCTGTGTGGAGCGGTTGTGGATGCGGTGGGCTGGGAGCAT 780
DB 814 TCCGAGGGGCTGGGCTCTCTCTCTGAGCGCTCTCGACCGGCTGGCAAGGACCAT 873
QY 781 CCGGCTCTGGCGGTGTAGCGCGCAGTGGCTCAACAGAGCGGTGCTGATGGGCTG 840
DB 874 CGGATCTCTGGCTGGTCCGGGGCACCGGCTCAACAGAGCGGCGCCAGAGCGGCTCT 933
QY 841 ACCGCGCCGAAACGGTCCGGCTCAGAGCGGGGTATCCGGCAGCGTTGGGAAACGGCGG 900
DB 934 ACCGCTCCGAAACGGGCGGTGCGAGCGCGGTATCCGACGTGCTCCGCGGCGACCGCG 993
QY 901 TTGCTCGGTGGCGGATGGATGGGTGGAGGGGAGCGGAGCGGAGCGGCGGCGGTAT 960
DB 994 CTCACGACCTCCGACGTGGACGTGTGAGGCGCCAGCGCACGGGTAGCGGACTCGGCGAC 1053
QY 961 CCGATCGAGGACAGGCGGTGTGCTCGCACGTAACGGGCGAG---CGGGCGGTGACAGGCGG 1017
DB 1054 CCGATCGAGGCGGAGGCGGCTCATCTCGCACGTAACGGGCGAGGCGGCTGACCGGCGAA 1113
QY 1018 CTGTGCTGGGGTCTCTGAAGTCCAAATCGGCGACACCATGGCTGCGCGGCTGGGT 1077
DB 1114 CTGCGCTCGGGTCTGTGAAGTCCAAATCGGCGACACCATGGCGCGGCTGGGTCTCC 1173
QY 1078 GGGGTATCAAGATGGTGAATGGGCTGGGAGGGGGTGTGGCGGAGCGTTCATGTG 1137
DB 1174 GGGGTATCAAGATGGTGAATGGGCGGATGGCGCACGGGCTCTGCGAAGAGCTCCACGTG 1233
QY 1138 GATTAACCGGTGCGCGAGGTGAGTGGTCCGCGGGGCGGTGCGGCTGCTGACGAGGCG 1197
DB 1234 GATTAACCGGTGCGCGAGGTGAGTGGTCCGCGGGGCGGTGCGGCTGCTGACGAGGCG 1293
QY 1198 GTGCGCTGGCGGAGGACCGCGGAGGGCGGTTGCGCGGCGGAGGTGCTGCTGGG 1257
DB 1294 ATGAGCTGGCGGAGCAAGGCGGAGCTGCGGAGGCGCGGCTCTCTCTTCGCG 1353
QY 1258 ATGCGGCGGACGAAATGGCGATGATTTTGAAGAGCGCGCGGCG 1302
DB 1354 GTGAGCGGAGCAAGCGGACGTGTGCTGAAGAGCGCGCGGCG 1398

RESULT 13
US-09-105-537-34
Sequence 34, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ. ID NOS: 43
SOFTWARE: PatsBIO for Windows Version 3.0
SEQ ID NO 34
LENGTH: 4689
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-34

Query Match 48.0%; Score 625.4; DB 4; Length 4689;
Best Local Similarity 68.6%; Pred. No. 2.8e-114;
Matches 895; Conservative 0; Mismatches 401; Indels 9; Gaps 2;

QY 1 GACCCGATCCGATCATCGGATGAGTGGCTTCCGGGCGGAGTCCGCTCCCGAAG 60
DB 103 GAGCGGATGCGATGAGGATGAGGCTGCGGCTGCGGCGGATGCGCTGCGCGAG 162
QY 61 GACCTGTGAGGCTGCGGCTCGGCGGAGAGCGCATCGGCGGCTTCCGACCGAGCG 120

DB 163 GACCTGTGAGGCTGCGGCTCGGCGGAGAGCGGATCTCGAGATTCCCGAGAGCGCG 222
QY 121 GATAGGCGCCAGGAAACAGGCTCACGCGCCAGAGACCGCGGAGCGTTCTATCG 180
DB 223 GGTGAGGAGTGTGAGGAGGCTGTACGACCCGACCGGAGCGGCTCGGAGAGACGTATCG 282
QY 181 CAGGAGGCGGAGTCTCTACGAGCGGCGGACCTTCAAGCGCGGCTTCTCGAATAGT 240
DB 283 CGGTGCGGAGGATCTCTACGAGCGGCGGAGGATTCAGCGGACCTTCTCGGATCTCG 342
QY 241 CCAGTGAAGGACCTGCGATGATTCGACGAGCGGCTGCTGTGAGAGACGTCTGAGAG 300
DB 343 CCGCGGAGGCGCTGCGCATGAGCCGAGAGGACGATGTCTTCACCAACCGCGTGGAG 402
QY 301 GGTTCGAGCGGCGGAGATGATCCGCTGTCCGATACCGGAGTCCGTAACGGGCTTTC 360
DB 403 GCGATGAGAGGCGGCGGATGACCCGACCGGCGCTGAAGGAGAGGCGCTCGGCTTTC 462
QY 361 GCGGCGGCGCTCTCTTGAATACGCGCGGATGAGACAGCGGCTGTCGAGGCGCG 420
DB 463 GTGCGGCGGCTGCGACACCGGCTACACCTCGGCGGAGACGACCG-----CGTGAAGT 516
QY 421 GCGGAGTGAAGGCGCACATCTCAACCGGTACACGCGGAGCGTCTGTGCGGCGGTATC 480
DB 517 CCGGAGTGAAGGCGCACATCTGATGACGCGGCGGCGGCTGTGCTGCGGCGGTATC 576
QY 481 GCGTACAGCTTCCGAGTGAAGGCGGCGGATCAACCGGTGACACGCGGAGTCTCGGCGAT 540
DB 577 GCGTACAGCTTCCGAGTGAAGGCGGCGGATCAACCGGTGACACGCGGAGTCTCGGCGAT 636
QY 541 CTGTGAGGCTGATCTGAGCGGCTGAGTGTGCGGCTGAGTGAAGTGAAGCTCGGCTG 600
DB 637 CTGTGAGGCTGATCTGAGCGGCTGAGTGTGCGGCTGAGTGAAGTGAAGTGAAGTGA 696
QY 601 GCGGCGGCGCTGCTGATGATTCACCTCGGAGATGTTTCAATGATTTCCCGGAGCG 660
DB 697 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 756
QY 661 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 757 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 816
QY 721 GCGGAGGCGGCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 817 GCGGAGGCGGCGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 876
QY 781 CCGGCTGCTGCGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 877 CCGGCTGCTGCGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
QY 841 ACCGCGCGGAAACGGTCCGCTCAGAGGCGGATCCCGGAGCGGCGGAGCGGCGG 900
DB 937 ACCGCTCCGAGAGGCGGCTCCGAGAGCGGATCCGAGAGCGGCGGCGGCGGCGG 996
QY 901 TTGCTCGGTGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 997 CTCGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1056
QY 961 CCGATGAGGACAGAGCGTGTCTCGCACGTAACG---GACGCGGCGGATGACAGCGG 1017
DB 1057 CCGATGAGGCGGAGCGGCTCATCGCACCTACCGGCGGAGGAGGAGGAGGAGGAG 1116
QY 1018 CTGTGCTGGGGTCTCTGAAGTCCAAATCGGCGCACCATGAGGCTGCGGCGGATG 1077
DB 1117 CTGAGGCTGGGGCTGTTGAAGTCCAAATCGGCGCACCATGAGGCGGCGGCGGATG 1176
QY 1078 GGGGTATCAAGATGGTGAATGGGCTGGGAGGGGGTGTTCGCGGAGCTTTCATGTG 1137
DB 1177 GGGGTATCAAGATGGTGAATGGGCTGGGAGGGGGTGTTCGCGGAGCTTTCATGTG 1236
QY 1138 GATTAACCGTCCGCGAGGTGAGTGGTCCGCGGGGCGGTTGCTGTGACGAGGCG 1197
DB 1237 GACGAGCGCTCGGAGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1296

QY 241 CCAAGTGAAGCACTGGGATGATCCGACAGCGGCTGCTGAGAGACCTCTGGAG 300
DB 5306 CCGCGCAGAGCCCTCGCATGAGACCCGACAGCGGCTCTCTCGAAGACTCTGGAG 5365
QY 301 GCGTTGAGGGGGGGAATCGATCCGCTGCGTACCGCGGCTCCGTAAGGCGCTTC 360
DB 5366 GCGGTGAGGAGACCGGGGATCGACCCGACCTCCCTGGGGGAGCGGAGGTGCGCTTC 5425
QY 361 GCGGGGCGCTCTCTCTGACTACGAGCCCGGATGAGACCGGCGTGTGAGAGGCGC 420
DB 5426 ACTGGGCGGATGACCCAGAGTACGGGCGAGCT-----GCGGAGCGGCGG 5473
QY 421 GCGGAGTGAAGGAGCCATCTTACCGGTACACGAGGAGCGTCTGTGCGGCGCTATC 480
DB 5474 GAAGGCTCGACCGCTACCTGCTGACCGGCAACAGGCGACGCTGATGTGGGCGCGCTC 5533
QY 481 GCGTACAGCTTGGGGCTGAGAGGCGGCGATCATCGGTGACACGGGGTCTCGGATCG 540
DB 5534 TCGTACACTCGGCTTGAAGGCGCGCGCTTGAACGCGCTGCTGCTGCTG 5593
QY 541 CTGCTGACGCTGATGAGCGGTGCAAGTCCGCTGCGGTGAGTGCAGCTCGCGCTG 600
DB 5594 CTGCTGCGCTTGAACCTCGCGGTGAGCGCTTGGCAAGGCGAGGCGTCAATGCGCTC 5853
QY 601 GCGGCGGCGCTCTGCTGATGTCACCCCTCGGATGTTCACTGAGTTCTCCCGGACGCG 660
DB 5654 GCGGCGGCGGTGCGCTGATGTCACCGCGGATGTTGCTGAGTTCAAGCGGCGAGCG 5713
QY 661 GCGCTGCTGCTGAGACGCGAGTGCAGAGCGTACTCGGCTGACAGCGGACCGCGCTG 720
DB 5714 GCGCTGCGGCGGAGCGCGGCTGAGAGCGCTTGCAGCGGCTGCGGAGCGGACCACTG 5773
QY 721 GCGGAGGCGCTGCGGATGCTGTTGAGAGCGGTTGCGGATGCGGTCGCGGCGAT 780
DB 5774 TCGGAGGCGGTGCGGCTCTCTCTGCTGAGCGCTTGGACCGCGCGGCGAGCGAC 5833
QY 781 CCGGCTGCTGCGGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 840
DB 5834 CAGGTCTCGGCTGCTGCGGCGGACGCGCTTGAACAGGACGCGCGGAGCGGCGCTC 5893
QY 841 AGCGCGCGGACGCTGCGGCTGACGAGCGGCTGATCCGCGAGCGCTTGGCGAACTGCGG 900
DB 5894 AGCGCTCGGACGCGGCTGCTGCGAGCGGCTGATCCGCGCGCGCTGCGGAGCGCGCG 9593
QY 901 TGTGCTGCGGATGAGTGTGAGAGGCGGACGCGGCGGACGAGCGCTGCGGTGAT 960
DB 5954 CTGACGACCTCGACGCTGAGCTGCTGAGAGCGACAGGACGCGGACTCGGCGAC 6013
QY 961 CCGATGAGGACAGGCGGCTGCTGCGCACGTAAGCGGCGG---CGGCGGCTGACAGCGG 1017
DB 6014 CCGATGAGGCGGCGGCGGCTGATCGCACCTTACGCGCGGCGGCTGACGAGGAGCGG 6073
QY 1018 CTGCTGCTGCGGCTCTGAAAGTCAACATCGGCGACACATGCGCTGCGCGGCTGCGG 1077
DB 6074 CTGCGGCTGCGGCTGTTGAAAGTCAACATCGGCGACACCGGCGCGGCGGCTGCTC 6133
QY 1078 GCGGCTGATCAAGTGTGATGCGTGTGCGGAGGCGGCTGCGCGGAGCTTGAAGTGTG 1137
DB 6134 GGTGTGATCAAGTGTGATGCGGAGTGCAGCGGCTGCTGCGGAGGCGGCTGCGAGCTC 6193
QY 1138 GATTAAGCTGCTGCGGAGTGAAGTGTGCGGCGGCGGCTGCGGCTGCGAGGAGG 1197
DB 6194 GACGAGCGCTCGGAGACAGATCGATGCTGCTGCGGCGGCTGCGGAACTCTTCAAGGAG 6253
QY 1198 GTGCGGCTGCGGAGGAGCGGCGGCGGCTGCGGCGGCGGAGTGTGCTGCTGCGG 1257
DB 6254 GTGCTGCTGCGGAGGAGGAGCGGCGGCGGCTGCGGCGGCGGCTGCTGCTGCGG 6313
QY 1258 ATCGGCGGAGCGAATGCGGATGATTTTGGAGAGGCGCGGCGG 1302
DB 6314 ATCGGCGGAGCGAATGCGGATGATTTTGGAGAGGCGCGGCTG 6358

RESULT 12
US-09-105-537-32
; Sequence 32, Application us/09105537A
; Patient No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Lau, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32

Query Match 48.7%; Score 634.2; DB 4; Length 11220;
Best Local Similarity 69.1%; Pred. No. 5,9e-116;
Matches 902; Conservative 0; Mismatches 388; Indels 15; Gaps 2;

QY 1 GACCGGATGCGCATCATCGGATGCGATCGGCTTCCCGGCGGAGTCCGCGGAG 60
DB 106 GAGCGGATGCGCATCATCGGATGCGATCGGCTTCCCGGCGGAGTCCGCGGAG 165
QY 61 GACCTGTGAGAGCTGCGCGGCTGCGGCGGAGAGCGGATCGGCGGCTTCCCGGAG 120
DB 166 GACCTGTGAGAGCTGCGCGGCTGCGGCGGAGAGCGGATCGGAGTCCCGGAGAG 225
QY 121 GATGCGGCGGAGAGAGAGTCAAGCGGCGGAGAGAGCGGAGCGGAGCGGCTTATCG 180
DB 226 GATGCGGAGAGTGAAGAGCGGCTGTAAGCGGAGCGGAGCGGAGCGGAGTGAAG 285
QY 181 GAGGAGGCGGAGTCTTCAAGAGCGGCGGAGCGGCTTCAAGCGGAGTGAAGTCA 240
DB 286 GCGGAGGCGGAGTCTTCAAGAGCGGCGGAGCGGAGTGAAGCGGAGTGAAGTCA 345
QY 241 CCAGTGAAGGCACTGAGCGATGATCGGAGCGGAGCGGAGTGAAGCGGAGTGAAG 300
DB 346 CCGCGGAGGCGGCTGCGGATGAGCGGAGCGGAGTGAAGCGGAGTGAAGCGGAG 405
QY 301 GCGTTGAGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 360
DB 406 GCGTTGAGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 465
QY 361 GCGGCGGCGGCTGCTGCGGAGTGAAGCGGAGCGGAGTGAAGCGGAGTGAAGTGA 420
DB 466 ACCGCGGATGATCAAGAGTGAAGCGGAGCGGAGTGAAGCGGAGTGAAGTGAAG 513
QY 421 GCGGAGTGAAGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 480
DB 514 GAGGCGGATGAGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 573
QY 481 GCGTACAGCTTGGGCTGAGAGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540
DB 574 GCGTACAGCTTGGGCTGAGAGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 633
QY 541 CTGCTGAGCGGATGATGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 600
DB 634 CTGCTGAGCGGATGATGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 693
QY 601 GCGGCGGCGGCTGCGGATGATGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660
DB 634 GCGGCGGCGGATGATGCGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 753
QY 661 GCGGCTGCTGAGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 720
DB 754 GCGGCTGCTGAGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 813

```

; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and plikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FaSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

```

Query Match 49.1%; Score 639; DB 4; Length 13842;

Best Local Similarity 69.3%; Pred. No. 76-117;

Matches 905; Conservative 0; Mismatches 385; Indels 15; Gaps 2;

```

QY 1 GACCCGATCGCATATCGGATGAGATGCGTTTCCGCGCGAGATCGGATCCCGAAG 60
DB 3325 GAGCCGATGCGATCGTCCGATGCGCTGCGCGCTGCGCGATGCGCTGCGCGAG 3384
QY 61 GACCTGTGGAGCTGCGCGCTGCGCGAGAGATCGATCGGCTTCCCAACGAGCG 120
DB 3385 GACCTGTGGAGCTGCGCGCTGCGCGAGAGATCGATCGGATGCGCTGCGAGACCG 3444
QY 121 GATGCGCCACGGAACAGGCTGACGCGCCAGGACCCGACGCGCGGACGTTCTACG 180
DB 3445 GGTGTGAGATGAGAGAGGAGGCTGACACCCGATCCGAGACACCCGACGATGACGTC 3504
QY 181 CAGGAGGCGGCTTCTTCAAGACGCGCGGCGCATTCGACCGCGCTTCTTCCGATCACT 240
DB 3505 CGCAGAGGCGGCTTCTTCAAGAGAGTCCGCGCTTCCGACCGCGCTTCTTCCGATCTCG 3564
QY 241 CACGATGAGGATCGGATGAGATCGGACAGAGCGGCTGCTGAGAGATCTTCCGAG 300
DB 3565 CCGCGAGAGCGCTTCCGATGAGACCGGACAGCGGCTGCTTCCGAAACCTTCCGAG 3624
QY 301 GCGTTGAGCGCGGAGATCGATCGGCTGCGTACGCGCGCTTCCGATGAGGCTTCC 360
DB 3625 GCGGTGAGAGACCGCGGATGAGACCGGATCGGCTTCCGAGAGAGGCTGCGGCTTCC 3684
QY 361 GCGGCGCGCTTCTTCAAGACGCGCGGCGCATTCGACCGCGCTTCTTCCGAGGCGCG 420
DB 3685 ACTGGGCGGATGAGACCGAGATGAGGCGGCGGCTTCCGAGAGGCGGCGGCTTCC 3732
QY 421 GCGGAGCTGAGAGGCGGATCTTCAAGAGAGGCGGAGAGGCGGCTTCTTCCGAGGCG 480
DB 3733 GAGGCGCTTCAAGAGGCGGATCTTCAAGAGAGGCGGAGAGGCGGCTTCTTCCGAGGCG 3792
QY 481 GCGTACGATCTTCCGCGGCGGAGAGGCGGAGGCGGATCGGATGAGAGGCGGCTTCC 540
DB 3793 TCGTACGATCTTCCGCGGCGGAGAGGCGGAGGCGGATCGGATGAGAGGCGGCTTCC 3852
QY 541 CTGATGAGCTGATCGGATGAGATCGGATGAGATCGGATGAGATCGGATGAGATCGG 600
DB 3853 CTGATGAGCTTCCGATGAGATCGGATGAGATCGGATGAGATCGGATGAGATCGG 3912
QY 601 GCGGCGGCGCTTCCGATGAGATCGGATGAGATCGGATGAGATCGGATGAGATCGG 660
DB 3913 GCGGCGGCGCTTCCGATGAGATCGGATGAGATCGGATGAGATCGGATGAGATCGG 3972
QY 661 GCGGCTGCGGATGAGAGGCGGATGAGAGGCGGATCGGATGAGAGGCGGATGAGAGG 720
DB 3973 GCGGCTGCGGATGAGAGGCGGATGAGAGGCGGATCGGATGAGAGGCGGATGAGAGG 4032
QY 721 GCGGAGGCGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 780
DB 4033 TCGAGAGGCGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 4092
QY 781 GCGGCTGCGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 840
DB 4093 CAGGCTTCCGCGGCTTCAAGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAGAGG 4152

```

```

QY 841 ACAGCGCGGAGAGGCTTCCGATGAGAGGCGGATGAGATCGGATGAGAGGCGGATGAG 900
DB 4153 ACAGCGCGGAGAGGCTTCCGATGAGAGGCGGATGAGATCGGATGAGAGGCGGATGAG 4212
QY 901 TTGTCGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 960
DB 4213 CTGAGAGCTTCCGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 4272
QY 961 CCGATGAGAGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 1017
DB 4273 CCGATGAGAGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 4332
QY 1018 CTGATGAGAGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 1077
DB 4333 CTGATGAGAGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 4392
QY 1078 GCGGCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1137
DB 4393 GGTGTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 4452
QY 1138 GATGAGCGGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 1197
DB 4453 GATGAGCGGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 4512
QY 1198 GTGCGGCGGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 1257
DB 4513 GTGCGGCGGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 4572
QY 1258 ATGCGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 1302
DB 4573 ATGCGCGGAGAGGCTTCCGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 4617

```

RESULT 11

US-09-105-537-5

Sequence 5, Application US/09105537A

Patent No. 6265202

GENERAL INFORMATION:

APPLICANT: Sherman, D. H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

TITLE OF INVENTION: DNA encoding methymycin and plikromycin

FILE REFERENCE: 600.438US1

CURRENT APPLICATION NUMBER: US/09/105,537A

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FaSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 36778

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-105-537-5

Query Match 49.1%; Score 639; DB 4; Length 36778;

Best Local Similarity 69.3%; Pred. No. 86-117;

Matches 905; Conservative 0; Mismatches 385; Indels 15; Gaps 2;

```

QY 1 GACCCGATCGCATATCGGATGAGATGCGTTTCCGCGCGAGATCGGATCCCGAAG 60
DB 5066 GAGCCGATGCGATCGTCCGATGAGATGCGCTTCCGCGCGAGATCGGATCCCGAAG 5125
QY 61 GACCTGTGGAGCTGCGCGCTTCCGCGCGAGAGATCGGATGAGATGAGATGAGATGAG 120
DB 5126 GACCTGTGGAGCTGCGCGCTTCCGCGCGAGAGATCGGATGAGATGAGATGAGATGAG 5185
QY 121 GATGCGCCACGGAACAGGCTGACGCGCCAGGACCCGACGCGGACGTTCTATCGG 180
DB 5186 GGTGTGAGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAGAGGCGGATGAG 5245
QY 181 CAGGAGGCGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 240
DB 5246 CCGAGAGGCGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 5305

```



```
Db 4156 CCGATTCAGAGCGAGGCTCTGATCCACCTTACGCGCCAGGCGCTGACGACGAGCCG 4215
QY 1018 CTGTGGCTGGAGGCTCTCTGAAGTCCACATCGAGGACACCAATGCTGCGGGGTGTGGGT 1077
Db 4216 CTGCGCTCGGGGTCTGTGAAGTCCAAATCATGCGGAGACCCAGGCGCGGCGGTCTCC 4275
QY 1078 GGGGTCTCAAGATGCTGATGGCGGTTCGGGAGGGGGGTGTTCGCCGGAAGCTTGATGTG 1137
Db 4276 GGTGTATCAAGATGTGTCCAGGCGATGCGCACGATCTGCTCCGAAGACGTGACAGTC 4335
QY 1138 GATAGCGCTCGCGCGAGGTGACTGTCTCGCGGGCGGGTGGCGGCTGTGACGAGAGCG 1197
Db 4336 GACGAGCGCTCGAGACAGATGACTGTGTGCTGTGCGCGGTGAATCTTCCACCGAGGCC 4395
QY 1198 GTGCGCTGCGCGGGAGACGCGGACAGGCGGTTCGCGGCGGAGAGTGTCTGTTCGGG 1257
Db 4396 GTGACTGCGCGGAGAGAGACGAGGCGGTCTGCGGCGCGCGCGCTCTCTTCGAG 4455
QY 1258 ATCGCGCGACGATGCGATGTGATTTTGAAGAGCGCGCGCG 1302
Db 4456 ATCAGCGGACCAATGCGATGTGTCTGGAAGAGCGCGCGTG 4500
```

```
RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1
```

```
Query Match 49.3%; Score 641.6; DB 4; Length 15872;
Best Local Similarity 70.1%; Pred. No. 2.2e-117;
Matches 914; Conservative 0; Mismatches 369; Indels 21; Gaps 3;
```

```
QY 1 GACCCGATCGCATTCATCGCATGAGCATGCCGTTTCCGGGCGGAGTCCGGTCCCGAAG 60
Db 8498 GATCCGATCGCATTCGATGCGCATGCGATGCGCTATCCGGGTGTGTGTGTGCGCGAG 8557
QY 61 GACCTTGGAGAGTGTGCGCGCTTGGGCGAGAGACGCAATGGGCGCTTCCCGACCGAG 120
Db 8558 GACCTTGGAGAGTGTGCGCGCTTGGGCGAGAGACGCAATGGGCGCTTCCCGACCGAG 8617
QY 121 GATGCGCCACGAGACAGAGGTACGCGCCAGGACCCGACGAGCGCGGACGTTCTATCCG 180
Db 8618 GGTGTGAGACTTGAAGAGCTTGAAGACCGGATCCGAGTCCGAGGAGGACCAAGTATCCG 8677
QY 181 CAGGAGAGCGGCTTCTTCAACAGACGCGGACCTTTCAGACCGCGGCTTCTTCCGAATCACT 240
Db 8678 CCGGAGAGCGGCTTCTTCAACAGACGCGGACCTTTCAGACCGCGGCTTCTTCCGAATCACT 8737
QY 241 CCACTGAGGCACTGCGCATGATTCGACAGACGCGGCTCTGTGTGAGAGCTTCGAGAG 300
Db 8738 CCGCGGAGAGCGCTTGTGTGATGAGCCGACAGCGGCTCTGTGTGAGAGCTTCGAGAG 8797
QY 301 GGTTCGAGAGCGGCGGAGATTCGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Db 8798 GGTTCGAGAGCGGCGGAGATTCGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8857
```

```
QY 361 GCGGCGCGCTCTCTTCCATGAGCGCCCGCATGTGAGACCGCGCTGTGAGAGCGCC 420
Db 8858 GTGGGCGCGCGCGACGCTCTGTAC-----GCTTCCGATCCCGGCTGTGTGCC 8905
QY 421 GCGGAGGTGAGAGGCGACATCTTCAACGAGTACAGGAGGAGCGGTCTGTGCGGCGATTC 480
Db 8906 GAGGAGCTCGAGAGGCTATCTGTGACGAGGAGCGGACGCGGATGTGTGTGTGTGTGTGT 8965
QY 481 GCTTACAGCTTCCGGCTGAGAGGCGGAGATCACCGGTGAGACAGGAGGCTTCCGAGATCG 540
Db 8966 TCTTACAGGCTGTGTGTGAGAGGACCGTCCATGACGCTGTGAGAGCGGCTGTCTCTCTG 9025
QY 541 CTGTGACGCTGATGTGCGGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
Db 9026 CTGTGTGCGCTGTGATGTGCGGTGACAGGCGCTGTGCGGACGCGGAGTGTGTGTGTGT 9085
QY 601 GCGGCGGCGGTCTCGGTTCATGTTCACCTTCGAGATTTATGAGATTTCCCGGACGCG 660
Db 9086 GCGGCGGCGGTGTGCGGTGTGATGTGCGGTTCGAGCGGCGGTGTGTGTGTGTGTGTGT 9145
QY 661 GGGCTGTGAGTGAACGAGAGTGAAGGCTACTGCGGTGACGCGGACGAGCGACCGGCTGG 720
Db 9146 GGGCTGTGAGTGAACGAGAGTGAAGGCTACTGCGGTGACGCGGACGAGCGACCGGCTGG 9205
QY 721 GCGGAGGCGGTGTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
Db 9206 GCGGAGGCGGTGTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9265
QY 781 CCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
Db 9266 ACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9325
QY 841 ACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Db 9326 ACAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9385
QY 901 TTGTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 960
Db 9386 TTGTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9445
QY 961 CCGATGAGAGCACAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1017
Db 9446 CCGATGAGAGCACAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9505
QY 1018 CTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1077
Db 9506 CTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9565
QY 1078 GGGGTCTATCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1137
Db 9566 GGGGTCTATCAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9625
QY 1138 GATTAAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1197
Db 9626 GATTAAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9685
QY 1198 GTGCGGTGCGGAGGAGACGCGGACGCGGCTTTCGCGGCGGAGTGTGTGTGTGTGTGT 1257
Db 9686 GTGCGGTGCGGAGGAGACGCGGACGCGGCTTTCGCGGCGGAGTGTGTGTGTGTGTGT 9739
QY 1258 ATCGCGCGACCAATGCGATGTGATTTTGAAGAGCGCGCGCG 1301
Db 9740 ATCGCGCGACCAATGCGATGTGATTTTGAAGAGCGCGCGCG 9783
```

```
RESULT 10
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
```

```

Db      50969 GGTCTGCTCCGGATGAGCGGTGCAAGTCTTCGCGCGGCGCGGATGACACCGGGGTG 51028
Qy      721 GCGAGAGGCGTCCGGATGCTGTTGTGTGAGCGGTTTCCGATCCGGTGGCGCTGGGGCAAT 780
Db      51029 GCGAGAGGCGTCCGGATGCTGTTGTGTGAGCGGTTTCCGATCCGGTGGCGCTGGGGCAAT 51088
Qy      781 CCGGCTCTGAGCGGTGTGACGCGGACGTGCGTCAACAGAGACCGTCTGTAATGGGCTG 840
Db      51089 GAGGTTCTGCGCGGTGTGCGGGGTACCGCGGTGACAGAGACCGCGCTGTAATGGGCTG 51148
Qy      841 ACGGCGCCGAAACCGTCCGGCTCAGAGCGGGTATCCGGACGCGTGGCGAACCGCGGG 900
Db      51149 ACTGCCCAATATGTCCTGTACAGCAAGAGGTATACCCAGGCACTCAAGAGTCCGG 51208
Qy      901 TTGTCTGCTGCGGATGTGATGTGTGTGAGAGGGGACAGGGACAGACGCTGGGTAT 960
Db      51209 CTGTCTGCTGCTCCACCTGTGATGCTGTGAGAGCGCATGGGACCGCGCTTGTGAT 51268
Qy      961 CCGATCGAGGACAGGCGTTGCTCGCACGTACG---GGCAGCGGGCGGGTGAACAGGCGG 1017
Db      51269 CCGATCGAGGCGGAGCGCTTGTGATGCTGTGAGAGCGCGGATGTGATCCCGTGGCGG 51328
Qy      1018 CTGTGCTGCGGCTCTGTGAGTCAACATCGGACACCAATGCTCCCGGGTGTGGT 1077
Db      51329 TTGTGCTGCGGCTGCGGTGAAGTGAATATGCTCAACACCGAGCGGCGGGGTGTGCT 51388
Qy      1078 GGGGTATCAAGTGTGATGTGTGTGCGGGAGGGGGTGTGCGCGGACGTTGCAATGTG 1137
Db      51389 GGGGTATCAAGTGTGATGTGTGTGCGGGAGGGGGTGTGCGCGGACGTTGCAACGTG 51448
Qy      1138 GATAGACCGTCCGCGAGGTGACTGTGCTCCGCGGGGCGGTGCGGCTGCTGACGAGGCG 1197
Db      51449 GACGAGCGCTCCGCGAGGTGACTGTGCTCCGCGGGGCGGTGCGGCTGCTGACGAGGCG 51508
Qy      1198 GTGCGCTGCGCGGAGCGCGGAGCGGGCTGTGCGCGGCGGAGTGTGCTGCTGCGG 1257
Db      51509 AGCCCTGCGGCT---CGACAGCGGCTGCTTCCGCGGGGCGGGCTGTGATGCTGCGG 51562
Qy      1258 ATGCGCGGACGAATCGGCTGTGATTTTGGAGAGGCGCGG 1299
Db      51563 ATCAGTGGACCAACGCGGACCTGATCTTGAACCAACCTCG 51604

RESULT 8
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506

```

```

; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
Query Match      49.3%; Score 642.2; DB 3; Length 38506;
Best Local Similarity 69.5%; Pred. No. 1.9e-117;
Matches 907; Conservative 0; Mismatches 383; Indels 15; Gaps 2;

Qy      1 GACCCGATCCGATATGCGGATGCGGATGCGGCTTCCCGGCGGATCCGGTCCCGGAAG 60
Db      3208 GAGCGGTGGCGATGTGCGGATGCGGATGCGGCTTCCCGGCGGATCCGGTCCCGGAAG 3267
Qy      61 GACCTGTGAGAGCTGCGCGCTCGGGCGGAGACGCGATCGGCGCTTCCCGACCGACGCG 120
Db      3268 GACCTGTGAGAGCTGCGCGCTCGGGCGGAGACGCGATCGGCGCTTCCCGACCGACGCG 3327
Qy      121 GGATGGCCACGGAACAGCTGACGCGGACGCGGACGCGGACGCGGACGCTTCTATCCG 180
Db      3328 GGCTGGAGACGTGAGAGGCGCTGACACCGGATCCGAGCACCCGCGACGCTGTAACGTC 3387
Qy      181 CAGGAGAGCGGCTTCTTCAAGACGCGGACGCTTCAAGCGCGGCTTCTTGGAAATCACT 240
Db      3388 GCGCAGGCGGCTTCTTCAAGACGCGGACGCTTCAAGCGCGGCTTCTTGGAAATCACT 3447
Qy      241 CACGCTGAGGCACTGCGATGATCGGACGAGCGGCTGCTGAGAGACGCTGAGAG 300
Db      3448 CCGCGCGAGGCGCTTCCGATGAGACCGGACGAGCGGCTTCTTGGAAATCACTGAGAG 3507
Qy      301 GCGTTCAGACGCGGCGGGAATGATCGCTGTGAGTACGCGGCTCCGTAACGGGCGTCTTC 360
Db      3508 GCGCTTCAGAGACGCGGAGATGACCGGACGCTTCCGCGGAGACGAGGCGGCGTCTTC 3567
Qy      361 GCGGCGCGCTTCTTCTTGAAGTACGCGGCGGCTGAGACCGGCGGCGGCGGCGGCG 420
Db      3568 ACTGGGCGGATGACCGAGAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3615
Qy      421 GCGGAGTGTGAGAGGCGGACATCTTCAACGCGTACAGGCGGAGCGGCTGCGGCGGATC 480
Db      3616 GAGGCGCTTCAAGGCTTCTGAGTACCGGACAGGCGGAGGCTGATGCGGCGGCGGCG 3675
Qy      481 GCGTACAGCTTGGGCTGGAAGGCGGCGGATGACCGGCGGAGTGTGAGTGTGCGGCGT 540
Db      3676 TCGTACACGCTGCGGCTTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3735
Qy      541 CTCTGAGCGCTGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 600
Db      3736 CTGTGCGCTTGCACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3795
Qy      601 GCGGCGGCGGCTGCGGATGCTGCAACCGTCCGCGGATGCTGAGTGTCTCCGCGAGCGC 660
Db      3796 GCGGCGGCGGCTGCGGATGCTGCAACCGTCCGCGGATGCTGAGTGTCTGAGTGTCTG 3855
Qy      661 GGGCTGTGCTGAGAGCGGAGTGAAGGCGTACTGCGGCTGACGCGGACGCGGCGGCTG 720
Db      3856 GGGCTGTGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3915
Qy      721 GCGGAGGCGGCTGCGGATGCTGTTGTGAGAGCGGCTGTGAGTGTGCGGCTGCGGCGG 780
Db      3916 TCGGAGGCGGCTGCGGCTTCTTCTGAGCGGCTGTGAGAGCGGCGGCGGCGGCGGCG 3975
Qy      781 CCGGCTGCTGCGGATGCTGAGCGGCGGAGTGAAGGCGTCAACAGAGCGGAGTGTGAGT 840
Db      3976 CAGGCTCTGCGGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4035
Qy      841 ACGGCGCCGAAACGCTCCGGCTCAGAGCGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCG 900
Db      4036 ACGGCTCCGAAACGCGGCGCTTCCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4095
Qy      901 TTGTCTGCTGCGGATGTGATGTGTGTGAGAGGGGACGCGGACGCGGACGACGCTGGGTAT 960
Db      4096 CTGACGACCTTCAAGTGTGATGTGTGTGAGAGGGGACGCGGACGCGGACGCGGACG 4155
Qy      961 CCGATCGAGGACAGGCGGCTTGTGCTGCGCACGTAACGGGCGAG---CGGGCGGGTGAACGGCGG 1017

```

```

QY 301 GCCTTGAGCGCGGCGGGAATGATCCGCTTCGCTGACCGCGGTCCTGTAACGGGCGCTTTC 360
DB 50618 ACCCTTGAGAGCGGCTGGGATGATCCGCTTCGCTGACCGGACCGGATCGCTTC 50677
QY 361 GCGGCGCGCCCTCTCTCTTCACTAAGCGCGCGGATGAGACACCGGCTGCTGGAGAGGCGCC 420
DB 50678 GCGGCGCTTATGATACAGAGCTATGGGCGCGGATTCATTAACAGAGACCGGAGGCGCTTC 50737
QY 421 GCGGAGCTGAGAGGCGCAATCTCAACCGGTACACAGGCGGACGCTCTGCGGCGGTATC 480
DB 50738 -----GAAGGGAACCTCGGGAAGGAGCATGCGGGAAGCGGTGCTGCGGT 50788
QY 481 GCTTACAGCTTCGGGCTGAGAGGCGCGGATACCGGTGACACCGGCGTCTCGGATC 540
DB 50789 GCGTATTCCTGTTGATTCAGAGGCTCTGCGGTACCGGTGATACCGGCTGCTGCTG 50848
QY 541 CTCGTACGCTGATCTGCGGTGCGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 600
DB 50849 TTGATGAGCTTACACTGCGGCTCAAGACCTGCGGCGCGGTGCGGTGCGGTGCGGT 50908
QY 601 GCGGCGCGGCTCTGCTGATATGTCACCTTCGCGGTATTCATGAGTTCTCCGCGAGCG 660
DB 50909 GCGGCGGAGGTACAGGTGATGTCAGCGCGGACGACGCTTCGTGAGTTCTCCGTAACG 50968
QY 661 GCGCTGCTGCTGAGAGCGGACAGTTCAGAGCGGTACTGCGCTGCGAGCGGACCGGCTG 720
DB 50969 GGTCTGCTGCTGAGAGCGGCTGAGAGCGGTGCGGCGGCGGCGGATGCGGCTG 51028
QY 721 GCGGAGGCGGCTGCGGATGCTGTTGAGAGCGGCTTCGCGGTGCGGTGCGGTGCGGTG 780
DB 51029 GCGGAGGCTGCGGCTGCTGTTGAGAGCGGCTTCGCGGTGCGGTGCGGTGCGGTGCG 51088
QY 781 CCGGCTGCTGCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTGCTGAGAGCGG 840
DB 51089 GAGGTTCTGCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTGCTGAGAGCGG 51148
QY 841 AGCGGCGCGGAGCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTGCTGAGAGCG 900
DB 51149 ACTGCGCGCAATGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTGCTGAGAG 51208
QY 901 TTGCTCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 960
DB 51209 CTGCTCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 51268
QY 961 CCGATCGAGGACAGGCTGCTGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 1017
DB 51269 CCGATCGAGGACAGGCTGCTGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 51328
QY 1018 CTGCTGCTGAGAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 1077
DB 51329 TTGCTGCTGAGAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 51388
QY 1078 GCGGCTGCTGAGAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 1137
DB 51389 GCGGCTGCTGAGAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 51448
QY 1138 GATTAAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 1197
DB 51449 GATTAAGCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 51508
QY 1198 GTGCGGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAG 1257
DB 51509 AGCGCTGAGAGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAG 51562
QY 1258 ATTCGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 1299
DB 51563 ATTCGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAG 51604

```

```

; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madhuri, Krishnamurthy
; APPLICANT: Treadway, Paul J
; APPLICANT: Turner, Jan R
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-1

```

```

Query Match      50.0%; Score 650.4; DB 4; Length 80161;
Best Local Similarity 70.5%; Pred. No. 5.2e-119;
Matches 918; Conservative 0; Mismatches 366; Indels 18; Gaps 3;

```

```

QY 1 GACCGGATGCGGATGATGAGAGCGGCTTCGCGGCGGAGTCCGCTCCCGGAG 60
DB 50318 GATCTTGTGCGATGCTGCGGATGAGAGCGGCTTCGCGGCGGAGTCCGCTCCCGGAG 50377
QY 61 GACTGTGAGAGCTGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGTCCCGGAGCGG 120
DB 50378 GAGCTTGTGCGGCTGCTGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGTCCCGGAGCGG 50437
QY 121 GATGCGGCGGAGAGAGCGGCTGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGAG 180
DB 50438 GATGCGGCGGAGAGAGCGGCTGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGAG 50497
QY 181 CAGGAGGCGGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCG 240
DB 50498 TGTGAGGCGGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCG 50557
QY 241 CCAGTGAAGAGAGCGGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGG 300
DB 50558 CCGGCTGAGAGGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCG 50617
QY 301 GCGTTGAGAGGCGGAGAGTATGATCGGCTTCGCGGCGGAGAGCGGCTTCGCGGAG 360
DB 50618 ACTTGAAGAGGCGGAGAGTATGATCGGCTTCGCGGCGGAGAGCGGCTTCGCGGAG 50677
QY 361 GCGGCGGCGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCG 420
DB 50678 GCGGCGGCTTCGCGGCTTCGCGGCGGAGAGCGGCTTCGCGGCGGAGAGCGGCTTCG 50737
QY 421 GCGGAGCTGAGAGGCGGATCTCAACCGGTACAGAGCGGCTTCGCGGCGGAGAGCGG 480
DB 50738 -----GAAGGGAACCTCGGGAAGGAGCATGCGGGAAGCGGTGCTGCGGT 50788
QY 481 GCTTACAGCTTCGGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 540
DB 50789 GCGTATTCCTGTTGATTCAGAGGCTCTGCGGTACAGAGCGGCTTCAGAGCGGCTTCAG 50848
QY 541 CTCGTACGCTGATCTGCGGTGCGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 600
DB 50849 TTGATGAGCTTACACTGCGGCTCAAGACCTGCGGCGGCTTCAGAGCGGCTTCAGAG 50908
QY 601 GCGGCGGCGCTGCGGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGG 660
DB 50909 GCGGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAG 50968
QY 661 GCGCTGCTGAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAGCGGCTTCAGAG 720

```

RESULT 7
 US-09-370-700-1
 ; Sequence 1, Application US/09370700

Db 31517 CCGGAGGCGGCTTCTCCACGACGCGGCGCATTCGACGCGGACCTTCTTCGGAGTTTCG 31576
Qy 241 CCACGTGAGGCACTGGCGATGAGTCCGACAGACGGCGCTGTGTGGAGACGTCTTCGGAG 300
Db 31577 CCGCGTGAAGCGCGGTGCGATGAGTCCGACAGACGGCGCTGTGTGGAGACGTCTTCGGAG 31636
Qy 301 GCGTTGAGAGCGGCGGGAATCGATCCGCTGTGGTACGCGGCGTCCGCTACGCGGCGTCTTC 360
Db 31637 GCGATCGAAGCGGCGCGGTATCGACCGCGAGAGACTCCCGCGACCGCGCGGCGGTAC 31696
Qy 361 GCGGCGCGCGCTCTCTTCGACTACGCGCGCGGTATGAGACCGCGCTGTTCGAGGCGCGC 420
Db 31697 GTGGGCGCGCTGAGACAGCGGCTACACCGGCGAGCGCGACCGCGCTCG----- 31744
Qy 421 GCGGAGCTGAGAGGCGACATCTCCACCGGTACACGCGGCGCGCTCTGTTCGGCGCGTATC 480
Db 31745 GCGGAGCTGAGAGGCGACATCTCCACCGGTACACGCGGCGCGCTCTGTTCGGCGCGTATC 31804
Qy 481 GCGTACAGCTTCCGCGTGGAGAGGCGCGCATACCGGTGACACGCGGCGTCTCGCATCG 540
Db 31805 GCGTACAGCTTCCGCGTGGAGAGGCGCGCATACCGGTGACACCGCGGCGTCTCGCATCG 31864
Qy 541 CTCGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 31865 CTCGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 31924
Qy 601 GCGGCGCGCGCTCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 31925 GCGGCGCGCGCTCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 31984
Qy 661 GCGGCTGCGGTGAGCGGACAGGTGCAAGCGCTGCTGCGCTGCGACGCGACGCGCGCTGCG 720
Db 31985 GCGGCTGCGGTGAGCGGACAGGTGCAAGCGCTGCTGCGCTGCGACGCGCGCTGCGCGCC 32044
Qy 721 GCGGAGGCGCTCGGAGTCTGTTGTTGAGACGCGTGTTCGAGTCCGCTGCGCTGCGGCGAT 780
Db 32045 GCGGAGGCGCTCGGAGTCTGTTGTTGAGACGCGTGTTCGAGTCCGCTGCGCTGCGGCGAT 32104
Qy 781 CCGGTGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 32105 CCGGTGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 32164
Qy 841 ACCGCGCGCGACGCTCGGCTCGAGAGCGGCTGATCCGCGAGCGGCTGCGGACGCGCGG 900
Db 32165 ACCGCGCGCGACGCTCGGCTCGAGAGCGGCTGATCCGCGAGCGGCTGCGGACGCGCGG 32224
Qy 901 TTGTCTGCTGCGGATGAGT 960
Db 32225 TTGTCTGCTGCGGATGAGT 32284
Qy 961 CCGATCGAGGCAAGCGCTGTGCTCGCACTACGGGC---AGCGCGCGGTTGACAGCGCG 1017
Db 32285 CCGATCGAGGCAAGCGCGCTGTGCTCGGAGACGTAATGCGGAGTGTGTGTGTGTGTGTGTGT 32344
Qy 1018 CTGTGCTGCGGCTCTGTAAGTCAACATCGGCGACACATGAGCTGCGCGGCGGTGTGGGT 1077
Db 32345 GTGTGCTGCGGCTCTGTAAGTCAACATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 32404
Qy 1078 GGGGTCTATCAAGT 1137
Db 32405 GGGGTCTATCAAGT 32464
Qy 1138 GATTAAGCGCTGCGCGAGT 1197
Db 32465 GATTAAGCGCTGCGCGAGT 32524
Qy 1198 GTGCG 1257
Db 32525 CCGGAGT 32578
Qy 1258 ATCGGCG 1291
Db 32579 GTGAGT 32612

RESULT 6
US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madhuri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Paul J.
APPLICANT: Turner, Jan R.
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1
Query Match 50.0%; Score 650.4; DB 3; Length 80161;
Best Local Similarity 70.5%; Pred. No. 5.2e-119;
Matches 918; Conservative 0; Mismatches 366; Indels 18; Gaps 3;
Qy 1 GACCGGATGCGCATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 60
Db 50318 GATCTTGTGTGATGTGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 50377
Qy 61 GACTGTGTGAGCTGTGCGGCTGTGCGGCGGAGACGCGCATGCGGCGCTTCCACGAGCGCG 120
Db 50378 GACTGTGTGAGCTGTGCGGCTGTGCGGCGGAGACGCGCATGCGGCGCTTCCACGATCGCG 50437
Qy 121 GATGCGCCACGGAACAGGTGACCGCCAGAGACCCCAAGCGACCGCGGACGTTCTATCGG 180
Db 50438 GACTGCGGATCTGTGCGGCGCTGTGACATCTGATCCCGATGCTGTGCGGACCTGTATGTGG 50497
Qy 181 CAGGAGGCGGCGGCTTCTTCAAGACGCGGCGCACTTCAAGCGCGGCTTCTTCCGAACTAGT 240
Db 50498 TGTGAGGCGGCGGCTTCTTCAAGACGCGGCGCACTTCAAGCGCGGCTTCTTCCGAACTAGT 50557
Qy 241 CACGTGAGGCACTGTGCGATGTGATCCGACGAGCGGCTGTGCTGTGAGAGCGTCTGTGGAG 300
Db 50558 CCGCGTGAAGCGGCTGTGCGATGTGATCCGACGAGCGGCTGTGCTGTGAGAGCGTCTGTGGAG 50617

Db 13647 CCGCGTGAAGGCGTTGGCGATGACCCGCGACAGCGGTTGCTGCTGAGACGCTCTGGAG 13706
QY 301 GCCGTTGAGCGGCGGAGATCGATCCGCTGTCGATCGCGGATCCCGTACGCGCGTCTTC 360
Db 13707 GCCCTGAGAGGCGCGGATCACTCTTCCAGTGGCGGGCACTCCGACCGGTTCTC 13766
QY 361 GCGGCGGCGCTCTCTTCACTACGAGCGCGGATGAGACACCGCGTGTGAGAGGCGCC 420
Db 13767 TTGGCGATGTGCAACGAGACTACGCGCGCGAGCGGCGGACGCGTCCGAG----- 13820
QY 421 GCGGACGTGAGGCGGCGCATCTTCAACCGGATACCGGCGGACCGCTCTGCGGCGGATC 480
Db 13821 -----CTGAGAGGCGTACCTGCTCAACCGGCTTCACTTCCAGCGTCTTGGGCGGTT 13874
QY 481 GCGTACAGTTCTGCGGCTGAGAGGCGCGGATGACCGTGAACACCGGCGGCTCTCGGATC 540
Db 13875 GCTTACAGCTTGGCTTGAAGGCGCGTGGATGAGTACGCGGTTGCTGCTGCG 13934
QY 541 CTGCTGACGCTGATCTGCGGTCGCACTGCGTGGGATGAGTGAACGCTTCCGCGCTG 600
Db 13935 TTGGTGGCGTTGATCTGCGGTCGAGGCGGTTGGCGGCGGATGTTGCTTGGCGTTC 13994
QY 601 GCGGCGGCGCTCTGCGGATGTCACCGCTGCGGATGTTCACTGAGTTCTCCGCGGCGC 660
Db 13995 GTGGGTGGGATGACGCGGATGCTGCGCGGTCGCTGACGCTTCCAGTCTGCGCGAG 14054
QY 661 GCGCTGCTGCTGAGACGCGGATGTCAGGCGGTCGCTGAGCGGCGGCGGCGGCTG 720
Db 14055 GCTTGTGCGTGAATGAGGCGGTCGAGGCGGCTTCCGCGGATGCTTGGTGGCT 14114
QY 721 GCGGAGGCGCTGCGGATGCTGTTGTTGAGAGCGGTTGTCGATGCGGCTGCGGCGAT 780
Db 14115 GCGGAGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 14174
QY 781 CCGGCTGCTGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 840
Db 14175 CCGGCTGTTGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 14234
QY 841 ACAGGCGCGGACGCTGCGGCTGCGGAGCGGATGCTGCGGAGGCTTGGGCGGCGG 900
Db 14235 GCGGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14294
QY 901 TTGTCGCTGCGGATGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 960
Db 14295 CTGGCTCTGCTGCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 14354
QY 961 CCGATCGAGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
Db 14355 CCGATCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14414
QY 1018 CTGTCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1077
Db 14415 GTTGTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14474
QY 1078 GGGGCTCATCAAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
Db 14475 GGGTGTGATGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14534
QY 1138 GATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197
Db 14535 GATTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14594
QY 1198 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1257
Db 14595 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14648
QY 1258 ATGCGCGGCGAGATGCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1302
Db 14649 ATGCGCGGCGAGATGCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 14693

RESULT 5
US-08-804-227C-1

Sequence 1, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Kosteck, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII (DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 52.0%; Score 676.4; DB 2; Length 43280;
Best Local Similarity 72.0%; Pred. No. 3.9e-124;
Matches 932; Conservative 0; Mismatches 341; Indels 21; Gaps 3;

QY 1 GACCCGATGCGCATCATGCGGATGCGATGCGGTTTCCGCGGCGAGTCCGTTCCCGAAG 60
Db 31337 GAGCGATGCGCGTGGTGGATGCGCTGCGCTACCGCGGAGGCGCAACACCCCGAA 31396
QY 61 GACCTGTGAGACTGCGCGCTTCCGCGGAGAGCGCATGCGGCGGTTCCCGACCGACCG 120
Db 31397 CAGTTCTGGGAAGTGTGACACCGGACCGACCGCGCGCGCATGCTTCCGACCGG 31456
QY 121 GGATGGCCACAGGACAGGCTGACGCGGACCGACCGACCGACCGACCGACCGACCGACCG 180
Db 31457 GGATGGACACCGACCGGCTGACGCGGACCGACCGGACCGGACCGGACCGACCGAC 31516
QY 181 CAGGAGGCGGCTTCTTCAACGCGCGGCGACCTTGAAGCGCGGCTTCTTGGAAATCACT 240

Best Local Similarity 72.6%; Pred. No. 1.1e-129; Matches 945; Conservative 0; Mismatches 342; Indels 15; Gaps 2;			
Qy	1	GACCCGATCGCATCATCGGCATGGCATCGCTTTCCGGGCGGAGTCGGTCCCGCAAG	60
Db	8270	GAGCGATCGGATCGTGGGATGGCTGCCGTTACCGGCGGTGTGGCGTCCCGGAC	8329
Qy	61	GACCTGTGGAGCTGGCGCTCGGGCGGAGACCCATCGGGCCGTTCCCCACACGACGC	120
Db	8330	GACCTGTGGACCTGGTGGCCGGTGAAGGGCACACGCTCTCCCGTCCCGGCCGACCGT	8389
Qy	121	GGATGGCCACGGAAACAGCTCAGCCACGAGACCCACGACGCCGCGACGTTCTATCCG	180
Db	8390	GGCTGGGACGTCGAGGGGCTGTACGACCCCGAGCGGGGTGCCGGGCAAGAGCTATGTA	8449
Qy	181	CAGGGAGCGGGTTCCTTACAGCGCGCGCACCTTCAGACGCCGCTCTTTCGGAATCAGT	240
Db	8450	CGGGAAGCGGGTTCCTGCTTCGCGGCCGAGTTCGACGCCGAGTCTTCGGGATATCG	8509
Qy	241	CCAGTGAGGCACCTGGCGATGGATCCGACGAGCGGCTGCTGTGGAGAGCTCTCGGGAG	300
Db	8510	CCGCGAGGCCACCGCCATGGACCCGACGACGCTGCTGTCTGGAGACGTCGTGGGAG	8569
Qy	301	CGCTTCAGACGGCGGGGAATCGATCCGCTGTTCGTTACGCGGTCCTGTAAGGGGCTTCTC	360
Db	8570	CGCTGTGAGCGGGCGCGCATCGTTCGGGACTCGCTGCGCGCACCCGAGCCGGTGTCTTC	8629
Qy	361	CGGGCGCCCTCTCTTCGATACGGCCCGGTATGACACCGCGTCGTGGAGGGCGCC	420
Db	8630	AGCGGCATCTCCAGAGGACTAC-----CGGACCCAGCTGGGGGACGCCGCC	8677
Qy	421	CGCGACGTGAGGGCCACATCCTCAACGGTACCAAGGGCAGCGTCTGTGGGGCGGTATC	480
Db	8678	GACACCTACGGCGGCGATGTCTCAACGGGACCCCTCGGCAGTGTGATCTCGGTCGGGTT	8737
Qy	481	GCCTACAGTTTCGGGCTGGAAAGGCCCGCGCATCAACCTTGACACGCGGTGCTCGGCATCG	540
Db	8738	GCCTATCGTTGGGTTTGGAGGGGCCCGCGCTGACCGTGGACACCGCGCTGTTCTGTCGTCG	8797
Qy	541	CTCGTAGCTGCATCTGGCGTCCAGTCCGTGCGGTCGGGTGAGTGACACGCTCGCGCTG	600
Db	8798	TTGTTGGCGTTGCATCTGGCGGTGCAGTCTGTTGCGCGGGGTGAGTGTGATGTCGGCGTTG	8857
Qy	601	GCCGGCGCGTCTCGGTCAATGTCACACCTCGGCATGTTCAATCGAGTTCCTCCCGCAGCGC	660
Db	8858	GCCGTTGGGTGACGGTGATGGCGACCGCGACGTTGTTCTGTGAGTTCCTCGCGCAGCGG	8917
Qy	661	GGGCTGTCGTGGACGCGAGGTCAAGGCGTACTCGGCTCGACCGCAGCGACCGGCTGG	720
Db	8918	GGGCTGGCGCGACGGCGCGGTGCAAGCGCTTCGCGAGGGTGGCGACGGCGCGCGGTGG	8977
Qy	721	GGCGAGGGCTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT	780
Db	8978	GCAGAGGTTGGGTTGTCTGTGTTGGAGCGGCTTTCCAGCGCGCGCCGCAACGGTTCAT	9037
Qy	781	CGGTTGTCGGGTGGTACCGGCGAGTGGCGTCAACAGGACGCTGCGTTCGAAATGGGCTG	840
Db	9038	CGGTTGTCGGGTGGTGGGGGAGTTCGGTCAATCAGAACGTTGCGAACATGGGCTG	9097
Qy	841	ACGCGCCGAAACGTTCCGGTTCAGGAGCGGGTGTATCCGGCAGGCGTTTGGCGAACCGGGG	900
Db	9098	ACGCGCCGAGTGTGTCGGCGCACACGCGGTGTATCCGTGAGGCGCTGGCTGATCGGGG	9157
Qy	901	TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGACCGGGAACGCGGACGACGCTGGGATGAT	960
Db	9158	CTGTGTCGCCCGACGTTGGATGTGGTGGAGCGCACGCTACGGGGACCGCGCTGGGTTGAT	9217
Qy	961	CCGATCGAGGACAGGGGTTGCTCGCCACGTCAGGGGACGGGCGGCTGACAGGCGCGCTG	1020
Db	9218	CCGATCGAGGCGGTTGCTGCTGGCCACGTAACGGCGGAGCGGGTTCGGCATTCGTTG	9277
Qy	1021	TGGCTGGGTTCTCTGAAGTCAACATCGGGCACACCATGCGCTGCGCGGCTGTGGGTGGG	1080

Db	9278	TGCTCGGGTCGTTGAAGTCGAACATCGGGCATCGCGAGCGCGCTCGGGGTGTGGGTGGT	9333
Qy	1081	GTCAATCAAGATGCTGATGGCGTTTCGGGGAGGGGGTGTTCGCCGGACGTTTGCATGTGGAT	1140
Db	9338	GTCAATCAAGTGTGTGAGGGGATCGGGCATGGGTCTTTCGGCGGACGCTGCATGTGGAT	9397
Qy	1141	AAGCCGTGCGCCGACGTGCACTGGTCCGGGGGGCGGTGCGGTGTGTGACGGAGGCGGTG	1200
Db	9398	GCGCCGTCTGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG	9457
Qy	1201	CCGTGCGCGGGGACCGCGCAGCGGGCGGTTCGGCGGGCGGGAGTGTCTGTTCCGGATC	1260
Db	9458	TCGTGCGCGCGG---CGGTGGAGCGGGTCCGCGGGCGCGCGGTTCGGCGTTCCGGGTG	9514
Qy	1261	GCGCGCACGAATCGCATGTGATTTTGGAGGAGCGCGCCGGCG	1302
Db	9515	AGCGGACCAACGCCCATGTGTCCTGGAGGAAGCGCCGGCG	9556
RESULT 3			
US-08-804-198-1			
; Sequence 1, Application US/08804198			
; Patent No. 5945320			
; GENERAL INFORMATION:			
; APPLICANT: Burgett, Stanley G.			
; APPLICANT: Kuhstoss, Stuart A.			
; APPLICANT: Rao, Nagaraja R.			
; APPLICANT: Richardson, Mark A.			
; APPLICANT: Rosteck, Paul R., Jr.			
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE			
; NUMBER OF SEQUENCES: 6			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: PAUL R. CANTRELL, 1138			
; STREET: LILLY CORPORATE CENTER			
; CITY: INDIANAPOLIS			
; STATE: IN			
; COUNTRY: USA			
; ZIP: 46285			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: Macintosh			
; OPERATING SYSTEM: Macintosh 7.0			
; SOFTWARE: Microsoft Word 5.1			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/804,198			
; FILING DATE:			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: CANTRELL, PAUL R.			
; REGISTRATION NUMBER: 36,470			
; REFERENCE/DOCKET NUMBER: P9113			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 317-276-3885			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 44377 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 350..14002			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 14046..20036			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 20110..31284			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: 31329..36071			
; FEATURE:			

Db 8255 GAGCCGATCGCGATGCGGCGATGCGTGCCTTACCCGGGCGGTGCGCGTCCCGGAC 8314
Qy 61 GACCTGTGGAGCTGCGGCGCTCGGGGAGAGCGCATCGGGCGGTTCGCCACCGACCGC 120
Db 8315 GACCTGTGGAGCTGCGGCGCTGAGCGGCGACACGCTCTCCCGGTTCGCCGCGGACCGT 8374
Qy 121 GGATGCCCAACAGAGCGGTACAGCCAGGAGCCCAACAGCGCGGCGACGTTCTATCG 180
Db 8375 GGTGTGGAGCTGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8434
Qy 181 CAGGAGGGGCTTCTTCAAGAGCGGCGCACTTCAGCGCGGCTTCTTCGGAATCACT 240
Db 8435 CGGAAAGGGGCTTCTGCGTTCGCGGCGGAGTTCAGCGCGAGTCTTTCGGGATATCG 8494
Qy 241 CCACGTGAGGCACTGCGGATGATCCGACGAGCAGCGGCTGCTGTGAGAGCTCTGGAG 300
Db 8495 CCGCGGAGGCCACGCGCCATGAGACCCGCGACGAGCGGTTCGTGTGAGAGCTGTGGGAG 8554
Qy 301 GCGTTCAGCGGCGGGAATCGATCGCTGCTCGGTACGCGGCTCCGTACGGGCGTCTTC 360
Db 8555 GCGCTGAGCGGCGCGCATCGTTCGGACTGCTGCGGGGACCCGAGCCGCTGCTTC 8614
Qy 361 GCGGGCGCTTCTTTCAGTACAGGCGCGGATGAGACACCGCGTGTGAGAGGCGCC 420
Db 8615 AGCGGCATCTCCAGCAGGACTAC-----GCAGCCAGCTGGGGAGCGCGCC 8662
Qy 421 GCGGAGCTGAGGCGGCACATCTCACCGGTACACCGGCGAGCGTCTGTGCGGCGTATC 480
Db 8663 GACACCTACGCGGGGATGTGCTCACGGGAGCCCTCGGCGATGTGATCTCCGCTCGGGT 8722
Qy 481 GCCTACAGCTTCGCGGTGGAAGGCGCGGATCACCGTGGACACGGGCTGCTCGGCATCG 540
Db 8723 GCCTATGCGTTGGGTTGGAGGCGCGGCTGACCGTGGACACGGCGTTCGTGCTCG 8782
Qy 541 CTGCTGAGCGTGCATCTGGCGTCCAGTCTGCTGCGGTGCGGTGAGTGCACGCTCGCGCTG 600
Db 8783 TTGGTGGCGTGCATCTGGCGGTGAGTCTGCTGCGGCGGCTGAGTGTGATCTCGCGGTTG 8842
Qy 601 GCGGCGGCGTCTCGGTGATGTCACCTCGGATGTTTATCGAGTCTCCCGGAGCGC 660
Db 8843 GCGGTTGGGTGACCGGTGATGGCGACCGCGTGTTCGTGAGTTCCTCGGCGAGCGG 8902
Qy 661 GCGCTGCTGCGTACCGCGAGTCAAGGCGTACTCGCTGCGAGCGGCGACCGGCTG 720
Db 8903 GCGCTGCGCGGACCGGCGGTCAAGGCTTCGCGGAGGCTGCGACGGGACCGGCTG 8962
Qy 721 GCGAGGCGTGGGATGCTGTTGTTGTTGAGCGGTTCGATCGGTGCGGTGCGGCTAT 780
Db 8963 GCGGAGGCTGCGGTGCTGCTGTTGGTGGAGCGGCTTTCGACGCGCGCGCAACGTCAT 9022
Qy 781 CCGGTGCTGCGGTGATCGCGGAGTGGGTCAACAGGACGCTCGTCAATGCGGCTG 840
Db 9023 CCGGTGCTGCGGTGATGCGGCGGAGTGGGTCAATCAGGACGCTGCGGAGCAATGGGCTG 9082
Qy 841 ACAGCGCGGAAACCGTCCGCTCAGGAGCGGCTGATCCGCGAGCGGTTCGCGAACGCGGG 900
Db 9083 ACAGCGCGGAGTGGTCCGCGCGAGAGCGGCTGATCCGTGAGCGCTGCGTATGCGGG 9142
Qy 901 TTGTCCGTGCGGATGATGTTGTTGTTGAGGCGACCGGACCGGCGACGCTGGGTGAT 960
Db 9143 CTGGTCCCGCGGACGCTGATGTTGTTGAGGCGCGTACGCGGAGCGGCTGGGTGAT 9202
Qy 961 CCGATCGAGGACAGCGGTGCTCGCACGTACGCGGCGCGGCGGCGGTGACAGCGCGCTG 1020
Db 9203 CCGATCGAGGCGGCTGCTGCTGGCGACGATCGGGCGGAGCGGCTGCGGATCCGTTG 9262
Qy 1021 TGGCTGGGTCTCTGAAGTCCAAATCGGCGACACCATGGCTCGCGGCGGTGGGTTGG 1080
Db 9263 TGGCTGGGTCTGTTGAAGTCGACATCGGCGATCGCGGCGGCTGCGGCTGGGTGGT 9322
Qy 1081 GTCATCAAGATGTGATGCGCTTTCGCGGAGGGGTGTTCGCGGAGCGTTGTCATGTGAT 1140

RESULT 2

US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
US-08-804-227C-7

Query Match

54.1%; Score 704.8; DB 2; Length 44377;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 56.4013 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gaccgcgcatcatcg...ttttggaggagcgccgcg 1302

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:
5: /cgn2_6/prodata/1/ina/PCUTUS_COMB.seq:
6: /cgn2_6/prodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	704.8	54.1	13987	2 US-08-804-227C-13	Sequence 13, Appl
2	704.8	54.1	44377	2 US-08-804-227C-7	Sequence 7, Appl
3	704.8	54.1	44377	2 US-08-804-198-1	Sequence 1, Appl
4	689	52.9	50937	4 US-09-428-517-1	Sequence 1, Appl
5	676.4	52.0	43280	2 US-08-804-227C-1	Sequence 1, Appl
6	650.4	50.0	80161	3 US-09-036-987A-1	Sequence 1, Appl
7	650.4	50.0	80161	4 US-09-370-700-1	Sequence 1, Appl
8	642.2	49.3	38506	3 US-09-320-878-19	Sequence 19, Appl
9	641.6	49.3	15872	4 US-09-105-537-1	Sequence 1, Appl
10	639	49.1	13842	4 US-09-105-537-30	Sequence 30, Appl
11	639	49.1	36778	4 US-09-105-537-5	Sequence 5, Appl
12	634.2	48.7	11220	4 US-09-105-537-32	Sequence 32, Appl
13	625.4	48.0	4689	4 US-09-105-537-34	Sequence 34, Appl
14	621.4	47.7	4403765	4 US-09-103-840A-2	Sequence 2, Appl
15	621.4	47.7	4411529	4 US-09-103-840A-1	Sequence 1, Appl
16	608.2	46.7	20235	3 US-07-642-734C-3	Sequence 3, Appl
17	608.2	46.7	20235	3 US-08-439-009A-3	Sequence 3, Appl
18	604	46.4	4411529	4 US-09-103-840A-1	Sequence 1, Appl
19	599.8	46.1	11219	1 US-07-642-734C-1	Sequence 1, Appl
20	599.8	46.1	11219	3 US-08-439-009A-1	Sequence 1, Appl
21	594.2	45.8	4403765	4 US-09-103-840A-2	Sequence 2, Appl
22	594.2	45.6	4041	4 US-09-105-537-36	Sequence 36, Appl
23	533.2	41.0	33529	4 US-09-144-085-3	Sequence 3, Appl
24	500.6	38.4	68750	3 US-09-335-403-1	Sequence 1, Appl
25	500.6	38.4	68750	4 US-09-568-102-1	Sequence 1, Appl
26	500.6	38.4	68750	4 US-09-567-969-1	Sequence 1, Appl
27	500.6	38.4	68750	4 US-09-568-480-1	Sequence 1, Appl

28	500.6	38.4	68750	4 US-09-568-486-1	Sequence 1, Appl
29	500.6	38.4	68750	4 US-09-568-472-1	Sequence 1, Appl
30	500.6	38.4	68750	4 US-09-567-899-1	Sequence 1, Appl
31	495.8	38.1	71989	4 US-09-443-501A-2	Sequence 2, Appl
32	488	37.5	28958	1 US-08-258-261B-6	Sequence 6, Appl
33	488	37.5	28958	1 US-08-456-837-6	Sequence 6, Appl
34	488	37.5	28958	1 US-08-457-342-6	Sequence 6, Appl
35	488	37.5	28958	1 US-08-457-646A-6	Sequence 6, Appl
36	488	37.5	28958	1 US-08-458-076A-6	Sequence 6, Appl
37	488	37.5	28958	1 US-08-764-233A-1	Sequence 4, Appl
38	488	37.5	28958	1 US-08-457-335A-6	Sequence 6, Appl
39	488	37.5	28958	1 US-08-729-214-6	Sequence 6, Appl
40	488	37.5	28958	3 US-09-028-934-6	Sequence 6, Appl
41	488	37.5	49377	1 US-08-764-233A-1	Sequence 1, Appl
42	460.4	35.4	1419	4 US-09-434-288-2	Sequence 2, Appl
43	411.2	31.6	1462	4 US-09-434-288-4	Sequence 4, Appl
44	338.4	26.0	1434	4 US-09-434-288-3	Sequence 3, Appl
45	321.6	24.7	751	3 US-09-010-809-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kunstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
US-08-804-227C-13

Query Match 54.1%; Score 704.8; DB 2; Length 13987;
Best Local Similarity 72.6%; Pred. No. 9.3e-130;
Matches 945; Conservative 0; Mismatches 342; Indels 15; Gaps 2;
OY 1 GACCGATCGGCATCATCGCATCGCGTTTCCCGGGCGGAGTCGGTCCCGGAAG 60

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 17, 2003, 11:26:45 ; Search time 1669.03 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gaccgagcgcacacacgcg.....tttcgagagagcgccgcgcg 1302

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_estum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gse:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pln:*
21: em_gse_vrt:*
22: em_gse_fun:*
23: em_gse_mam:*
24: em_gse_mus:*
25: em_gse_ocher:*
26: em_gse_pro:*
27: em_gse_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	8.5	1161	14	BM912505 AGENCOURT
2	110.6	8.5	1138	17	AG152606 Pan trogl
3	110.2	8.5	1201	14	B0063878 AGENCOURT
4	109.4	8.4	1243	14	BM913931 AGENCOURT
5	108	8.3	931	13	B1416575 haap001xg
6	108	8.3	1037	17	AG146864 Pan trogl

Result No.	Score	Query Match	Length	DB ID	Description
7	107.8	8.3	1089	14	B0930965 AGENCOURT
8	107.8	8.3	1288	14	BM809948 AGENCOURT
9	107.2	8.2	888	17	A2189327 SP_1014_A
10	107.2	8.2	918	17	A0895329 HS_4832_A
11	107	8.2	662	9	A1514012 CH27052_5
12	106.8	8.2	863	13	B1416699 B1416650
13	105.8	8.1	1015	13	B0939440 AGENCOURT
14	105.8	8.1	1089	14	B0939440 AGENCOURT
15	105	8.1	1197	13	B1416470 B0884996
16	104.8	8.0	1289	14	B0884996 B1416685
17	104.6	8.0	927	13	B1416685 B0890606
18	104.2	8.0	939	14	B0890606 B1416899
19	104.2	8.0	1222	17	B1416899 ENT0M477F
20	104	8.0	1137	17	AC078502 BM914619
21	104	8.0	104	14	BM914619 AL572700
22	103.8	8.0	855	9	AL572700 BE034822
23	103.8	8.0	1006	10	BE034822 ML04H02_M
24	103.8	8.0	1021	14	BQ441712 AGENCOURT
25	103.8	8.0	1089	13	BM468944 AGENCOURT
26	103.4	7.9	952	12	BC441341 GA_Ea001
27	103.2	7.9	1127	14	B0718629 AGENCOURT
28	103	7.9	960	14	B0429353 AGENCOURT
29	103	7.9	1005	17	BH146810 ENT267TR
30	102.8	7.9	670	17	AG126228 BM091001
31	102.4	7.9	625	13	BM091001 1g21a10.y
32	102.4	7.9	688	12	BC786336 SEAUMC006
33	102.4	7.9	798	13	B1416695 haap001xg
34	101.8	7.8	897	13	B1416517 haap001xd
35	101.8	7.8	1478	14	BM911600 AGENCOURT
36	101.6	7.8	877	17	AQ788449 HS_3130_A
37	101.4	7.8	1184	14	BM911719 AGENCOURT
38	101.2	7.8	1301	14	B0067141 AGENCOURT
39	101	7.8	1090	17	AG128386 AG128386 Pan trogl
40	101	7.8	1303	14	BM908878 AGENCOURT
41	100.8	7.7	715	9	AL570062 AL570062
42	100.8	7.7	817	17	A2196702 SP_1032_B
43	100.8	7.7	897	17	A2186654 SP_1007_B
44	100.8	7.7	924	10	BE455162 HVSMBh009
45	100.6	7.7	884	17	AQ782679 HS_3183_B

ALIGNMENTS

RESULT 1
LOCUS BM912505 1161 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6612272 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474051
5', mRNA sequence.
ACCESSION BM912505
VERSION BM912505.1 GI:19362884
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1161)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTP/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CH1986 row: k column: 12
High quality sequence etop: 218.
Location/Qualifiers 1. 1161

FEATURES
source

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:5474051"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC library."
BASE COUNT      53 a      844 c      107 g      88 t      69 others
ORIGIN

```

```

Query Match      8.5%; Score 111; DB 14; Length 1161;
Best Local Similarity 43.6%; Pred. No. 4.9e-10;
Matches 351; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

498 GGAAGGCGCGGATACCGGTGACACGGGCTGCTGGCATGCTGTGACGCTGATCT 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1061 GGAAGAGGNGANGAGANNAGGGAAGGGGGGNGGNGGGGGGGGGGGGGGGGGG 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
558 GGCCTGCCAGTCGCTCGGTCGGGTGATGACGCTCGGCTGGCCGCGCTCTCGGT 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1001 NCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
618 CATGTCACCTTCGCGATTTTCAGAGTTTCTCCCGCAGCGGGCTGTGCGTGACGG 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
941 GGGGNNNGGGGGGGGGGGGNNNGGGGGGGGGGGGGGGGGGGGGGGNNNGGGGG 882
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 CAGGTGCAGAGGGGTACTCGGCTGCAGCCGACCGGCTGGCGGAGAGGCTCGGAT 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
881 GNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 822
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 GCTGTTGTGAGCGGTTGTCGATTCGGATCGGCTGCGCTGAGCTCTGCGGTGAT 797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 762
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
798 AGCGGCGACGTGCGGTCAACGAGACGGTGCCTGGAATGGCTGACGCGCCGACACG 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
761 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858 GCGTACGAGCGGGGTGATCCGCGACGCCGTTGGCGAGACGGGGGCTTCTCCGTG 917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 642
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
918 GATGTGTGAGAGGGGACAGGACGGGACGATGAGTGAATCCGATCGAGGACAGGC 977
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
641 NCGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 582
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
978 GTTGCTCCGACATGACGGGACGGGCGCGGTGACAGCCGCTGTGGCTGCTCTGA 1037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1038 GTTCAACATCGGGCACACATGCTGCCCGGGTGTGGGTTCATCAATGATGAT 1097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
521 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1098 GCGCTTTCGAGAGGGGGTGTGCGCGGACGTTGATGTGATTAAGCCCTCCCGCAG 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
461 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1158 GACATGTCGCGGAGGGGCGGTGCTGTGAACGAGCGGTGCGCTGGCGCGGAGACGC 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
401 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1218 GGCAGAGGGGTTGCGGGCGGGGAGTGTCTTTCGGGATCGGCGGACGAAATGCGCA 1277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY      1278 TGTGATTTTGAGAGCGCCGCGC 1302
DB      281 GGGGGGGGGGGGGGGGGGGGGGGG 257

```

```

RESULT 2
AG152606
LOCUS
DEFINITION
Pan troglodytes DNA, clone: RP43-016B03.TU, genomic survey
ACCESSION
AG152606
VERSION
AG152606.1 GI:16682284
KEYWORDS
GSS.
SOURCE
Pan troglodytes male lymphocytes DNA, clone_11b:RP43-43 Chimpanzee
Male BAC library clone:RP43-016B03.TU.
Pan troglodytes

```

ORGANISM

```

REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Torok,Y., Matanabe,H. and Sakaki,Y.
BAC end sequences of Library RP43-43
2 (bases 1 to 1138)
Unpublished
JOURNAL
AUTHORS
Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-Chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbee@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RP43-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS

```

COMMENT

```

Sequencing: TU
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1..1138
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-016B03.TU"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP43-43 Chimpanzee Male BAC Library"
BASE COUNT      38 a      869 g      38 t      184 others
ORIGIN

```

FEATURES

source

```

Query Match      8.5%; Score 110.6; DB 17; Length 1138;
Best Local Similarity 38.9%; Pred. No. 5.8e-10;
Matches 389; Conservative 0; Mismatches 610; Indels 0; Gaps 0;

304 TTTCAGCGGGGGAATCGATCCGCTGCTGACGCGGCTCCGTAACGGGCTCTTCGCG 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 TTGTGAGGAGCGTGGNNNNNCGTGGAGTGTGGGNNNNNNNNNNNNNNAAAAAGAGG 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 GCGCCCTCTCTTCGACTACGCGCCCGGTATGACACCGCGTGTGAGAGGCGCGCG 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 GGGGGNNNNNNNNNNNNNNNNNNNNNNNGGNGGCGGNNNGGGGGGGGGGGGGGG 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
424 GACGTGAGGCGCATCTTCAACCGGTACACGCGGACAGCGTCTGTGCGGCGGTATGCC 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 GGGGGGGGGGGGNNGNGGNNNGGNNNGGNNNNNNNNNGGNGGCGGGGGGGGNN 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 TACAGCTTCGGGCTGGAAGGCGCGCGATCACCGTGTGACACGCGGCTGTCCGATCGCTC 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 NCGGGGNNNGGGGGGNNNNNNNNNGGGGGGGGGGGGGGGGGGGGGGGGGGG 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
544 GTGACGCTCATCTGCGGTGCACTGCGTGTGAGTGCACGCTGCGCTGCGC 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: WGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Plate: L1CM2097 row: 9 column: 13
High quality sequence stop: 152.

FEATURES
source location/Qualifiers
1..1201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5925540"
/clone_1lb="NIH MGC 99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB1; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT	23 a	41 c	763 g	15 t	359 others
ORIGIN					
Query Match	8.5%, Score 110.2; DB 14; Length 1201;				
Best Local Similarity	32.0%; Pred. No. 6.8e-10;				
Matches	283; Conservative 0; Mismatches 601; Indels 0; Gaps 0;				
QY	419	CCGCGGACGTGAGAGGCCACATCTCTACCGGTACACGGGCGAGCGTCTCTGCGGCCGTA	478		
DB	89	CCGAGTGTGTAGAGAGCGCGAGTGCAGGAGGAGGAGGCGGCTNGGCTGGGGGCGGN	148		
QY	479	TGCGCTACAGCTTGGGGCTGGAAGGGCGCGGATCACCGTGACACGGGCGTCTCGCAT	538		
DB	149	NNNNNNNNNGNNNNNNNNNNNNNNNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	208		
QY	539	CGCTCGTGCACGCTCATCTGCGCTGCCAGTCTGTCGGTCCGGTGAAGTGCACGCTCGGC	598		
DB	209	NN	268		
QY	599	TGCGCGGCGCGCTCTCGGTATGTCCACCTCGGCATGTTTCATGAGTTCTCCGCGACG	658		
DB	269	NN	328		
QY	659	GCGGGCTCTCGGTGAGACGGCAGGTGCAAGCGTACTCGGCTGCAGCCGACCGGCGCT	718		
DB	329	NN	388		
QY	719	GGGGCGAGAGCGTGGGATGTGTTGTGTGAGACGGTGTCCGATGCGGTGCGCTGGGCG	778		
DB	389	GGGAGNNNGGGNN	448		
QY	779	ATCGGGTGTGTCGCGTGTGTAACGCGGACGTGCGATCAACAGACGCTGCGTGAATGGC	838		
DB	449	GGGAGGGGNNNGGGNNNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	508		
QY	839	TGACGGCGCCGCAACGGTTCGGCTACGAGACGGGTGATCTCCGACAGCGTTGGCGAA	898		
DB	509	GGNGGGGNNNNGGGGGGGGGNNNNNNNNNNNNNNNGGGGNNNGGGGGGGGGGGGGG	568		
QY	899	GGTGTTCGCTGGCGGATGTGTAATGTGTAAGGGGACCGGACCGGACGACGCTGGGGT	958		
DB	569	GGNGGG	628		
QY	959	ATCCGATCGAGCACAGCGCTTGTCTGCGCACTGACCGGACGCGGCGGTGACAGCGCG	1018		
DB	629	GGGGGNNNGGG	688		
QY	1019	TGTGGCTGGGGTCTCTGAATGTCACCATCGGGCACACATGTGCTGCCGGGTGTGGGT	1078		
DB	689	GGNGGG	748		

[illegible]

Db	245	AGGGCAGGTCTCTGAGACCGGCGCCCTCTGGACGACAGGCGTACAGTACTCGTCTC	304
OY	531	CTCGGCATCGCTCGTGAACGTGCATCTGGCGTGCACGTGCTGCTCGTTCGGGTGATGCAC	590
Db	305	TTTTCAGGGGAGAAAGGGCTCCCCATGAGGAGAGGGGCGACAGGGGAGAGGGGGTGTATGAGGN	364
OY	591	GCTGCGCCTGGCCCGCGCGCTGTGTCGTATCTCCACCTCGGCATGTTCAATCGAGTTCTC	650
Db	365	MNNGNNNNNNNNNGNNGNNGNNGGAGGNNNGNNGNNGNNGNNGNNGGAGGAGGAGGNNGN	424
OY	651	CCGCGACAGCGCGAGCTGTCTCGTGAACGCGCAGGTGCAGAGCGTACTCGCTCGACGCCGACG	710
Db	425	NGGGGGGGGGGGGGGGGGGNNCGNNGNNGGAGNNGGAGGGGGGGGGGGGGGGGAGNAGNCGG	484
OY	711	CACCGCCTGGGGCGAGGGCGTCCGGGATGCTGTTGTGAGACGCGTTGTTCGATGCGGTGCG	770
Db	485	GGNCGGGGGGGGGGGGGGGGGGAGGAGNAGGAGGGGGGGGGGGGGGGGGGGGGGAGGNNGN	544
OY	771	GCTGGGGCATGCGGTGCTGCGCGGTGTACGCGGCAAGTGCCTCAACACAGACGCTGCGTC	830
Db	545	NGNAGGG	604
OY	831	GAATGGCGCTGACGCGCGCCGAAACGGTCCGGCTCAGAGACGCGGTGATCCGACGCGGTTGGC	890
Db	605	GGGT	664
OY	891	GAACGCGGGGTTGTCCGTGCGGATGTGATGTGTGAGAGGAGCACGGACGAGCACGAC	950
Db	665	TTGG	724
OY	951	GCTGGGTGATCCATCGAGGACAGGCGGTTCTCTCCACGTACCGGCGACGCGGCTCGTGA	1010
Db	725	GG	784
OY	1011	CAGGCGCGTGTGCGTGGGGGTCTTGAAGTCCMAATCGGGGACACCAATGCGCTGCCGCGGG	1070
Db	785	GG	844
OY	1071	TGTGGGTGGGGTCAATCAAGATGTGATGCGCTTGGCGGAGAGGGGTGTTGCCGCGACGTT	1130
Db	845	GG	904
OY	1131	GCATGTGGATTAAGCCGCTGCCGACAGGTGAGACTGTCCCGGGGGCGGTGCGGCTGTGAC	1190
Db	905	GG	964
OY	1191	GGAGGGCGGTGCGGTGGCCGGGGGAGCGGGGACGGCGGGGTGGCGCGCGGGCGGAGTGTGCTC	1250
Db	965	GG	1024
OY	1251	GTTCCGGATCGGCGG 1265	
Db	1025	GGGGGGGGGGGGGG 1039	
RESULT 5			
BI416575	931 bp	mRNA	linear EST 15-AUG-2001
LOCUS	haap001xg15f	Heterobasidion annosum -	Scots pine infection stage
DEFINITION	subcloning cDNA library (haap)	Pinus sylvestris/Heterobasidion	
ACCESSION	annosum cDNA clone haap001xg15f,	mRNA sequence.	
VERSION	BI416575		
KEYWORDS	BI416575.1	GI:15187598	
SOURCE			
ORGANISM	Pinus sylvestris/Heterobasidion annosum.		
REFERENCE	Pinus sylvestris/Heterobasidion annosum		
AUTHORS	Eukaryotica; mixed EST libraries.		
TITLE	1 (bases 1 to 931)		
	Asleebu,F.O., Nhalikova,T., Choi,W., Stenlid,J. and Dean,R.A.		
	Expressed sequence tags of randomly selected cDNA clones from the		
	interaction of the root rot fungus (Heterobasidion annosum) with		
	seedling roots of Scots pine (Pinus sylvestris)		

[illegible]

RESULT 7					
LOCUS	BQ930965				
DEFINITION	BQ930965	1089 bp	mRNA	linear	EST 20-AUG-2002
	AGENCOURT	8931577	NIH_MGC_40	Homo sapiens	CDNA clone IMAGE:6483877
	5', mRNA	sequence.			

ACCESSION	BQ930965
VERSION	BQ930965.1
KEYWORDS	GT:22345996
	EST.

SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

1 (bases 1 to 1089)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lln.gov>
plate: L16CM269 row: c column: 14
High quality sequence start: 28
High quality sequence stop:180.

```
FEATURES location/Qualifiers
source 1. .1089
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:648387"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/ncbi="organ: prostate; Vector: pOTM7; Site: 1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACACAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: This is a NIH_MGC library."

```

ORIGIN

Query Match	8.3%;	Score 107.8;	DB 14;	Length 1089;
Best Local Similarity	45.8%;	Pred. No. 1.8e-09;		
Matches 295;	Conservative 0;	Mismatches 349;	Indels 0;	Gaps 0;

653 GGCAGCGCGGGCTGTCCGTGACGGCAGGTGCAAGGCGTACTCCGGTGCAGCCGACGGCA 712

[illegible]

713 CCGGCTGGGGCGAGGCGCTCGGATGCTGTGGTGAGCGTTGTGGATGCGGTGCGGC 712

442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

773 TGGGCGATCGGGTCTGCGCGTGTACCCGCGAGTCCGCTCAACGAGACGGTGCCTCGA 832

[illegible]

833 ATGGGCTGACGGCCGCCACGGGCTCCAGGAGCGGGGATGATCCGGCAGAGCGTTCGCGA 892

893 ACCGCGGTTTGTCCGTGCGGATGTGCAATGGTGTGAAGGGGCA CGGGACGGGCA CGAGCC 952

D_b

622 GGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGNNNCGAGCGGGCGG 681

953 TGGGTGATCCGATCGAGGCACAGCGTTGCTCCCCACGCTACGGGCAAGCGGCGGTGACA 1012

[illegible]

1013 GAGCCGTGTGGTGGGGCTCTGAAGTCCAAACATCGGGCACACCATGGCTTGCCGGGGGTG 1072

[illegible]

QY 1073 TGGTGGGTCATCAGATGTGATGGCGTTGCCGAGGGGTGTGCCCGAGCTTGC 1132

[illegible]

1133 ATGTGGATTAAGCCGTCGCCGACAGTGAAC TGATCCGCCGGGGGGGTGGTGGGCTGCTGACGG 1192

[illegible][illegible][illegible]

Db

RESULT 8
BM809948

LOCUS	1288 bp	mRNA	linear	EST 05-MAR-2002
BM805948				
DEFINITION	AGENCOURT 6581186 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454872			

ACCESSION	3 / mRNA sequence.
VERSION	BM809948
BM809948	1
GI	19126771

VERSION	DATE	BY
KEYWORDS		
SOURCE		
human,		
EST.		

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

REFERENCE
1 (bases 1 to 1288)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
REMARKS 11-13-2001

COMMENT
JOURNAL
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: robsrb@earthlink.net

Email: cgadub-1@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Plate: LCM1950 row: 1 column: 09
 High quality sequence start: 7
 High quality sequence stop: 357.
 Location/Qualifiers

FEATURES

SOURCE

1.1288
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5454872"
 /clone_lib="NIH_MGC_98"
 /issue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library"

BASE COUNT
 75 a 131 c 783 g 62 t 237 others

ORIGIN

Query Match 8.3%; Score 107.8; DB 14; Length 1288;
 Best Local Similarity 39.4%; Pred. No. 1.8e-09;
 Matches 256; Conservative 0; Mismatches 393; Indels 0; Gaps 0;

654 GCACGCCGGCTCTCGGTGACCGACAGTCAGCGCTACTCGCTGACCCGACCGAC 713
 414 GGNNN 473
 714 CGGCTGGGGGAGGGGGTGGGATGCTGTTGGTGAAGCGGTTTCGATCCGGTGGCGCT 773
 474 GGGGGGGGNNNG 533
 774 GGGGATCGGGTCTGCGGTGTAACCGGCACTGCGTCAACAGGACGGTCTCGAA 833
 534 GGGGGGNNNGG 593
 834 TGGGCTGACCGCCCGACCGTCTCGGCTCAGAGCGGGTGTATCCGGACGGCTTGGCGAA 893
 594 GGGGGGNNNGG 653
 894 CGGGGGTCTCTCGGTGCGGATGTGTGTGAGAGGGCAAGGACGGGACGACGCT 953
 654 GGGNNNGGGNNNG 713
 954 GGGTGTATCCGATCGAGGACAGGCGTTTCTCGCCAGTACGGGACGGGCGGTGACAG 1013
 714 GGGGGGNNNG 773
 1014 GCCGTGTGCTGGGCTCTCTGAATGCCAATCGGACCAATGCGTTCGCCGGGTGT 1073
 774 GGGGNNNGGGGAGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 833
 1074 GGGTGGGCTCATCAATGTGTATGCGCTTGGGAGAGGGGTGTTCGCCGACGTTGCA 1133
 834 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 893
 1134 TGTGATTAAGCCCTCCCGCAGGTGACCTGTCCGCGGGGCGGTGCTGTGACGA 1193
 894 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 953
 1194 GCGGTGCTCGTGGCCGCGGAGACGCGGACGGGCGGTTGCGGCGGAGTGTCTGCTT 1253
 954 GGGGGGNNNG 1013
 1254 CGGGATTCGGCGGACGAATGCCGATGTTTGGAGAGAGCCGCCGGCG 1302
 1014 NGNNGGGGGCGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1062

RESULT 9
 AZ189327
 LOCUS
 DEFINITION
 SP_1014_A1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin
 , sperm genomic BAC library Strongylocentrotus purpuratus genomic
 clone Plate=1014 Col=9 Row=X, DNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT: Cameron, RA, Davidson, EH, Hood, L

Division of Biology 156-29

California Institute of Technology

Pasadena California 91125, USA

Tel: (626) 395-8421

Fax: (626) 795-3047

Email: acameron@caltech.edu

Plate: 1014 row: K column: 9

Seq primer: T7

Class: BAC ends

High quality sequence stop: 888.

Location/Qualifiers

1..888

/organism="Strongylocentrotus purpuratus"

/db_xref="taxon:7668"

/clone="Plate=1014 Col=9 Row=X"

/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"

/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"

BASE COUNT

52 a 37 c 606 g 83 t 110 others

ORIGIN

Query Match 8.2%; Score 107.2; DB 17; Length 888;
 Best Local Similarity 44.4%; Pred. No. 2.3e-09;
 Matches 274; Conservative 0; Mismatches 343; Indels 0; Gaps 0;

653 GGCAGCGCGGCTCTCGGTGTAACCGGACAGTCAAGCGTCTGACCGGACGCA 712
 272 GGGGGGNNNG 331
 713 CGGCTGGGGGAGGGCGTGGGATGCTGTTGTGAGCGGTTTCTCGATTCGGTGGCGC 772
 332 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 391
 773 TGGGATCGGGTGTGCGGTGTAACCGGACAGTCAAGCGGACGCTCTGCA 832
 392 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 451
 833 ATGGGCTGACCGGCGGCAACGCTCCGCTCAGAGCGGGTGTATCCGGACGGCTTGGCA 892
 452 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 511
 893 ACGCGGGGTTGCTCGTGGGAGATGTGTGAGAGGGGACGCGGACGGGACGACCG 952
 512 GGGGGGGGGGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 571
 953 TGGGTGATCCGATCGAGGACAGGCGTTTGTCCGACGTAAGGACGCGGCGGTGACA 1012

[illegible]

RESULT 10	
AQ895329	
LOCUS	AQ895329 918 bp DNA linear GSS 10-NOV-1999
DEFINITION	HS 4832 A2 H01 T7A CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=4832 Col=2 Row=O, DNA sequence.

VERSION	AQ895329.1	GI:6351435
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 918)

Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

TITLE	Sequence-tagged connectors: A sequence approach to mapping and screening the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3518
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com)
BAC and Web Server: <http://www.htec.washington.edu>
Plate: 4832 row: O column: 2
Seq primer: T7
Class: BAC ends
High quality sequence stop: 918.

FEATURES	Location/Qualifiers
source	1. .918

BASE COUNT	ORIGIN
16 a	93 c 718 g 60 t 31 others

Query Match	8.2%;	Score 107.2;	DB 17;	Length 918;
Best Local Similarity	49.1%;	Pred. No. 2.3e-09;		
Matches 301, Conservative	0;	Mismatches 307;	Indels 5;	Gaps 1

653 GGACGCCCGGCGTGTGGTGGACGGAGGTGCAAGCGCTGCTGGCTGCACGCCACGGCA 712

Db	305	GCGGGGCGCGCGCGGGGCCGAGNAGGGGGGGGGGGGGGGGGCCGCGGGCGGGGGG	364
OY	713	CCGCCTGGGCGCAGAGCGCTCGGGATGCTGTGGTGAACCGTTGTCCGATGCCGTGCGGC	772
Db	365	GCGGGGGCGGGGGGGGGGGGGCGGGCGGGGGGGGGGGGGGGGGGCGNGGGGGGGGG	424
OY	773	TGGGGCATTCGGAGTGCTGGCGGTGTGAGCGGCAAGTCGCTCAACAAGAACGGTGTCTCGA	832
Db	425	GGGGGGGGGGGGNAGGGGGGGGGGGGGCGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGG	484
OY	833	ATGGCGTCAAGCGCGCCGCAACGGTCCGGCTCAGAGACCGGTGATTCCGCGACGCTTGGCGA	892
Db	485	GGGGGGGGGGGGNAGGG	544
OY	893	ACGCGGGGATTGTCCTGTGCGCGGATGTGGATGTGGTAGGGGAGCACGGGACGAGCAGCGC	952
Db	545	GGGGGGGGGGGGGCGAGNGGGGGGGGGGGGGGGGGCGGGCGGGGGCGGGGGGGGGGGGGGG	604
OY	953	TGGGTGATCCGATCAGAGCACAGCGTTCGCAACGTACCGGCMAGCGGACCGGTGACA	1012
Db	605	GGGGGGGCGGGGGCGCGGG	664
OY	1013	GGCGGCTGTGGCTTGGGGGCTCTGTAAGTCAAACAATCGGGGACACACATGGCTGCGCGGGTGT	1072
Db	665	GCGCGGGGGGGCGGGGGGGGGGGGGCGGC-----GGGGGGGGGGGGGGGGGGGGGGGGGG	719
OY	1073	TGGGTGGGTCATCAGATGATGATGGCGTTGCGGGAGGGGGGTGTGCCCGCGACGTTTC	1132
Db	720	GGGGGCGCGGGGGCGGCGCGGG	779
OY	1133	ATGNGAMTAAGCCGCTCGCCGACGGTGAAGTGTCCGCGGGGGCGGATCGGCTGCTGACGG	1192
Db	780	GGG	839
OY	1193	AGCGGATGCCGTGGCGCGGGGGGACGCGGCAAGGGCGGTTGCGGCGGGCGGAGTGTGTCGT	1252
Db	840	GCGGGGGGGGGCGCGGGGGGGGGCGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	899
OY	1253	TCCGGATCCGCGCG 1265	
Db	900	GGGCGGGCGGGGG 912	

RESULT	11
A1514012	
LOCUS	
DEFINITION	662 bp mRNA linear EST 23-APR-2001
ACCESSION	GH27052.5prime GH Drosophila melanogaster head POT2 Drosophila
VERSION	melanogaster cDNA clone GH27052.5 similar to BCdNA:GH07625.
KEYWORDS	F88H00035323 located on: 2L 23D1-23D1; 04/10/2001, mRNA sequence.
SOURCE	A1514012.1 GI:4418074
ORGANISM	ESt,
	fruit fly,
	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neuroptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Epiphytoidae; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 662)
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
TITLE	Lewis,S.S. and Rubin,G.M.
JOURNAL	BDBP/HMMT Drosophila EST Project
COMMENT	Unpublished (2001)
Other_ESTS:	GH27052.3prime

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003581: arm:2L [2747932,3054367]
estimated-cyto:23B3-23D3: 04/10/2001
plate: GH.270 row: E column: 4

High quality sequence stop: 542
POLYA=No.

FEATURES
source location/Qualifiers
1. 662
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_id="GH27052"
/sex="male and female"
/dev stage="adult"
/lab_note="DH5 - alpha"
/note="Organ: head; Vector: POT2; Site: 1: EcoRI; Site 2: XhoI; Sited fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 143 a 185 c 185 g 149 t
ORIGIN

Query Match 8.2%; Score 107; DB 9; Length 662;
Best Local Similarity 50.7%; Pred. No. 2.5e-09;
Matches 316; Conservative 0; Mismatches 295; Indels 12; Gaps 2;

529 TGCTGGGCAATCGCTCGTGAACGCTGATCGGCGGCAAGTCGCTGGGCTGGGTGTCG 588
3 TGTCTCAAGTCTCTGTAACGCTTGTGAACAGGCTTTCTCCGATATGCGCAAGAGGTC 62
589 ACAGCTGCGGCTGAGCGGCGGCGCTCTCGATCATGCCACCTCGGCAATGTTTCATGAGTTC 648
63 GACAAAGCCCTGGTCCCTGAGAGCTGAGACTGATCTCTCAAGCCACATGTCGTCGACGTTTC 122
649 TCCCGGCAAGCGGCGGCTGTCGCTGGAACGCGAGGTGCAAGCGGCTACTCGGCTGACCGGAC 708
123 AAGCGACTGAATATGTTGAGCCCGGACGCGAGCTGCAAGCGCTCATGATGCTGCAAT 182
709 GGCACCGGCTGGGCGGAGCGGCGGCTGCGGATCGTGTGGAGGCTTGTCCGATATGCGGTC 768
183 GGTATAGCTCCGTTCCATGATGATGTCGCTCTCTCGAGCGCACCTCTGACGCAAG 242
769 CGAGCTGGGCGATCGGCTGCTGCGGCTGATACGCGGAGTGCAGGCTCAACGAGCGGTGCG 828
243 CGGTGTATAGCTTCATCTCTCAATATGTCGACCAATATGATGTTTCAAGAA----- 295
829 TCGAATGGGCTGAACGCGGCGGCAAGCGGCTCGGCTGAGAGCGGCTGATTCGCGAGCGGCTTG 888
296 --GACAGGCGATCATATACCTTATTTGGCAAGATGCAAAATGCGCTGATTCGCGAGACTAC 353
889 GCGAAGCGGCGGCTTGTCCGTGCGGATGTGATGTGTGAGGGGCAACGCGGAGCGGCGACG 948
354 GAGGAATATGCTTAAACCCCGCGATGTGTTTACGTGAGGCAACGCGTATCCGGAACC 413
949 ACAGCTGGGCTGATCGATTCGAGGCAACGCGGCTGCTGCGCACGTAACGAGGCGGCGGCT 1008
414 AAGGTGGCGATCCCGAGAGGTGAAGTCTATCATGACGATCTTCTTGCAA--GACCGT 470
1009 GACAGGCGGCTGTGCTGGGCTGCTGAAGTCCAAATCGGCGCAACATGAGGCTGCGCG 1068
471 ACAGCCCTCTGTGATCGGATCGGTCAAGTCCAAATGCGTACATCGGAACCCGCTTC 530
1069 GGTGTGGGTGGGCTCATCAAGATGTGATGCGCTTGGAGGAGGGGCTTGGCCGCGAGC 1128
531 GGTGTGTGCTCTGTGCGCAAGATTCGATGCGCATGAGAGAGGCGCTCATTTCCGGTAAAC 590
1129 TTGCAATGTGATTAAGCGGTCGCC 1151
591 TTGCACTACAAACAGCCGAACC 613

RESULT 12
BI416699 863 bp mRNA linear EST 15-AUG-2001
LOCUS
DEFINITION haep001xp01f Heterobasidion annosum - Scots pine infection stage
Subtraction cDNA library (haep) Pinus sylvestris/Heterobasidion
annosum cDNA clone haep001xp01f, mRNA sequence.
ACCESSION BI416699

VERSION BI416699.1 GI:15187722
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
REFERENCE Eukaryote, mixed EST libraries.
1 (bases 1 to 863)
AUTHORS Aislegbu, F.O., Nanaikova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Aislegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Aislegbu@mykopat.slu.se
Seq primer: T7 primer.
Location/Qualifiers

FEATURES
source 1. 863
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:159015"
/clone_id="haep001xp01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep)"
/dev stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site: 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (PP5)."
BASE COUNT 12 a 21 c 700 g 12 t 118 others
ORIGIN

Query Match 8.2%; Score 106.8; DB 13; Length 863;
Best Local Similarity 40.4%; Pred. No. 2.7e-09;
Matches 327; Conservative 0; Mismatches 482; Indels 0; Gaps 0;

494 GCGTGAAGGCGCGGCGATCAACCTGAGACACGCGGCTGCTCGGCAATCGCTGACGCTTCG 553
1 GCGCAACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 60
554 ATCTGCGCGTGCAGTGCCTGCGGCTGAGTGAACGCTCGCGTGCAGCGGCGGCTCT 613
61 AATGTGANNCCGGGCTGCTGCTAAGTCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
614 CGGTATGTCACCTCGGCAATGTTCAATGAGTTCCTCCGCGAGCGCGGCTGTGCGTGG 673
121 NGNN 180
674 ACCGCAAGTGTCAAGGCGCTACTGCGCTGACAGCGGACCGGCTGCGGCGGCGGCTGCG 733
181 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
734 GGATGCTGTGTGAGCGGCTTGTGCGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 793
241 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
794 TGTACGCGGCGAGTGCAGTCAACAGAGCGTGTGCTGATGAGGCTGACGCGGCGGCGGAG 853
301 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
854 GTCCGCTCAAGAGCGGCGGCTGCGGCGGCGGCTTGTGCGGCGGCGGCTTGTGCGGCGG 913
361 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
914 ATGTGATGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 973
421 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
974 AGCGTGTGCTGCCACGTAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1033

[illegible]

RESULT 13	LOCUS	DEFINITION
BI416650	1015 bp	mRNA
BI416650	linear	EST 15-AUG-2001
haep001xm01f	Heterobasidion annosum - Scots pine infection stage	

ACCESSION	BI416650
VERSION	BI416650.1
	GI:15187673

ORGANISM Pinus sylvestris/Heterobasidium annosum
Eukaryota; mixed EST libraries.
1/bases 1,401,917

AUTHORS Astiebu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

interaction of the root rot fungus (*Heterobasidion annosum*) with seedling roots of Scots pine (*Pinus sylvestris*)

JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu . . .

Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Astegren@mykopat.slu.se
Seq primer: T7 primer.

```
FEATURES location/Qualifiers
source 1. .1015
```

```

/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="haap001xm01f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtractions cDNA library (haap)"
/dev_stage="Seeding roots of scots pine were infected for
6 days with H. annosum"
/notes="vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FPS)."

```

Query Match	8.1%	Score 105.8	DB 13	Length 1015
Best Local Similarity	47.4%	Pred. No. 4e-09		
Matches 308; Conservative	0;	Mismatches 342;	Indels 0;	Gaps 0;

Oy 653 GGAGCGCGGCTGTTCGGTGCACGGCAGTGCAAGCGGTACTTCGGCTGCAGCCGACGCA 712

Dd 178 GGGAGGGGGGGGGGCGGCGGGGCGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 237

[illegible]

RESULT 14					
BQ939440					
LOCUS	BQ939440	1089 bp	mRNA	linear	EST 21-AUG-200
DEFINITION	AGENEOTD_8929085 NIH_MGC_40 Homo sapiens cDNA Clone IMAGE:6484345				
	5', mRNA sequence.				

ACCESSION	BO939440
VERSION	BO939440.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 1089)
AUTHORS	NIH-MGC http://mgc.ncbi.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCD/DNP
cDNA library Preparation: Rubin Laboratory
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLCM2670 row: 9 column: 02
High quality sequence stop: 327.

```
FEATURES
source
location/Qualifiers
1. .1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6484345"
```

```

/clone_1lb="NIH_MGC_40"
/feature_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pORF7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

```

BASE COUNT      73 a      59 c      712 g      81 t      164 others
ORIGIN
Query Match      8.1%; Score 105.8; DB 14; Length 1089;
Best Local Similarity 38.6%; Pred. No. 4e-09;
Matches 281; Conservative 0; Mismatches 447; Indels 0; Gaps 0;
QY 575 GGTGGGTGATGTCAGCTGCGCTGGCGGCGGCTCTCGTCAATGTCACCTCGGCA 634
    |||
DB 312 GGTAGAGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 371
QY 635 TGTTCATCGAGTTCTCCCGGACGGGGCTGTCGTGACGCGACAGTCCAGCGCTACT 694
DB 372 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGNGN 431
QY 695 CGGCTGACCGCAGCAGCAGCGCTGGGCGAGGCGCTCGGATGCTGTTGTCGACGGT 754
    |||
DB 432 NCGGCGNNGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 491
QY 755 TGTTCGATCGGTCGGCTGGGGCATCGGGTGTGCGCGGTGATCGCGGCAATGCGGTCA 814
    |||
DB 492 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 551
QY 815 ACCAGACCGTGTGTCGATGCGCTGACCGGCCGACGCTCCGCTCAGACGCGGTGA 874
    |||
DB 552 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 611
QY 875 TCCGCGAGCGTTGTCGGAACGCGGGGTTGTCGTGCGGATGTCATGTCGAGGGGC 934
    |||
DB 612 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 671
QY 935 ACGGAGCGGACGACGCTGGGTGATCGATCGAGGACACAGCGTTGCTCGCCACGTACG 994
    |||
DB 672 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 731
QY 995 GGCAGCGGGCCGTGACAGGCCGCTGTGGGTCTCTGAAGTCAACATGCGGACCA 1054
    |||
DB 732 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 791
QY 1055 CCATGCTGCGCGGGGTGTGGGTGGGTCATCAAGATGATGCGGTTGCGGGAGGGG 1114
    |||
DB 792 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGNGN 851
QY 1115 TGTTCGCGGACGTTGATGTCGATGATGATGATGATGATGATGATGATGATGAT 1174
    |||
DB 852 GGGGGGNNGGGNNNNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 911
QY 1175 CGGTGCGGCTGTCGACGAGGCGGTCCTGTCGCGGGGACGCGGACGCGGTCGCGC 1234
    |||
DB 912 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 971
QY 1235 GGGCGGAGATGTCGTTCCGGGATCGGCGGACGATGCGCATGTCATTTGGAGAGAG 1294
    |||
DB 972 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1031
QY 1295 CGCGGCG 1302
DB 1032 GGGGGGGGG 1039

```

RESULT 15
BI416470

```

LOCUS      BI416470      1197 bp      mRNA      linear      EST 15-AUG-2001
DEFINITION      haap001xa01f Heterobasidion annosum - Scots pine infection stage
                  subtraction cDNA library (haap) Pinus sylvestris/Heterobasidion
                  annosum cDNA clone haap01xa01f, mRNA sequence.
ACCESSION      BI416470
VERSION        BI416470.1 GI:15187493
KEYWORDS
SOURCE
ORGANISM      Pinus sylvestris/Heterobasidion annosum.
                  Pinus sylvestris/Heterobasidion annosum
                  Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 1197)
Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
Unpublished (2001)
JOURNAL
COMMENT
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@kpat.slu.se
Seq primer: 77 primer.
FEATURES
source
1..1197
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="haap01xa01f"
/clone_1lb="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (haap)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."
BASE COUNT      20 a      28 c      1036 g      28 t      85 others
ORIGIN
Query Match      8.1%; Score 105; DB 13; Length 1197;
Best Local Similarity 45.2%; Pred. No. 5.4e-09;
Matches 291; Conservative 0; Mismatches 353; Indels 0; Gaps 0;
QY 653 GGCACGCGCGCTGTCGATGACGCGAGTCGACGCGTCCGCTGACGCGACGCA 712
    |||
DB 355 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 414
QY 713 CCGGCTGGGGCGAGGCGCTCGGATGCTGTTGTCGAGCGTTGTCGATGCGTGGCG 772
    |||
DB 415 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 474
QY 773 TGGGCGATCGGTCGTCGCGGTGTAACGGCGCATGTCGCTCAACGACGTCCTCA 832
    |||
DB 475 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 534
QY 833 ATGGGCTGACGCGCGCAACGCTCCGGCTCAGAGACGGGTGATCCGGGACGGGTTGGCA 892
    |||
DB 535 GNNNGGNNGGGNNGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 594
QY 893 ACGCGGGGTTGTCGTGCGGATGTCGATGTCGTGAGAGGACGCGGACGCGACGCGC 952
    |||
DB 595 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 654
QY 953 TGGGTGATCCGATCGAGGACAGCGGTTGCTGCGACGTAACGGGACGCGGCGGTGACA 1012
    |||
DB 655 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 714
QY 1013 GCGCGCTGTCGTCGAGTCTCGAAGTCCAAATGCGGACACACATGCTGCGGGGNG 1072
    |||
DB 715 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 774
QY 1073 TGGGTGGGTCATCAAGATGTCGATGCGGTCGCGGAGGGGCTGTTGCCGCGAGTTGC 1132

```

[illegible]

Search completed: June 18, 2003, 01:00:11
Job time : 1674.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1786.55 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068

Perfect score: 921
Sequence: 1 gcttcgcttcctccgcgaca.....acaacgacacccgcgcttc 921

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280.

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_scs:*

12: gb_ey:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pac:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_scs:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pin:*

35: em_hcg_rtd:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_ey:*

39: em_hcgo_hum:*

40: em_hcgo_mus:*

41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921	100.0	12381	6 AX006889	AX006889 Sequence
2	921	100.0	30690	6 E38020	E38020 Avermectin
3	921	100.0	64957	1 AB032367	AB032367 Streptomycin
4	919.4	99.8	31422	6 E38021	E38021 Avermectin
5	919.4	99.8	64957	1 AB032367	AB032367 Streptomycin
6	677.2	73.5	11096	1 AF275943	AF275943 Streptomycin
7	402.4	43.7	65140	6 AX211705	AX211705 Sequence
8	402.4	43.7	123580	1 AF263912	AF263912 Streptomycin
9	402.4	43.7	125401	6 AX211739	AX211739 Sequence
10	401.4	43.6	104326	1 AB070940	AB070940 Streptomycin
11	401	43.5	113193	1 AF357202	AF357202 Streptomycin
12	400.8	43.5	113193	1 AX211706	AX211706 Sequence
13	385.4	41.8	20235	1 SERERYAB	M63677 S. erythraea
14	385.4	41.8	20235	6 AR049368	AR049368 Sequence
15	385.4	41.8	20235	6 AR095529	AR095529 Sequence
16	385.4	41.8	20444	1 SEERYAB	X62569 S. erythraea
17	381.6	41.4	104326	1 AB070940	AB070940 Streptomycin
18	375.6	40.8	84985	1 SNA278573	AJ278573 Streptomycin
19	373.8	40.6	24568	1 AF220951	AF220951 Streptomycin
20	373.8	40.6	50937	6 AR159871	AR159871 Sequence
21	371.2	40.3	11480	1 STMPK3J0RF	L09654 Streptomycin
22	362.8	39.4	41097	1 AF016585	AF016585 Streptomycin
23	355	38.5	78210	1 AB070949	AB070949 Streptomycin
24	351	38.1	53784	1 AMM223012	AJ223012 Amycolato
25	351	38.1	53789	6 A69720	A69720 Sequence 3
26	351	38.1	90445	1 AR040570	AR040570 Amycolato
27	349.6	38.0	7788	6 AX089456	AX089456 Sequence
28	349.6	38.0	11219	1 SERERYAB	M63676 Saccharopol
29	349.6	38.0	11219	6 AR049367	AR049367 Sequence
30	349.6	38.0	11219	6 AR095528	AR095528 Sequence
31	349.6	38.0	50000	6 AX089416	AX089416 Sequence
32	349.6	38.0	50000	6 AX089420	AX089420 Sequence
33	348	37.8	80161	1 AY007564	AY007564 Saccharop
34	348	37.8	80161	6 AR165018	AR165018 Sequence
35	347.4	37.7	47981	1 AF263245	AF263245 Micromono
36	347.4	37.7	47981	6 AX112026	AX112026 Sequence
37	345.6	37.5	3185	1 AF411573	AF411576 Actinomad
38	345	37.5	28958	6 AR044578	AR044578 Sequence
39	345	37.5	28958	6 I47768	I47768 Sequence 6
40	345	37.5	28958	6 I50958	I50958 Sequence 6
41	345	37.5	28958	6 I63356	I63356 Sequence 6
42	345	37.5	28958	6 I70387	I70387 Sequence 6
43	345	37.5	28958	6 I85639	I85639 Sequence 6
44	345	37.5	28958	6 I88045	I88045 Sequence 4
45	345	37.5	28958	6 I90322	I90322 Sequence 6

ALIGNMENTS

RESULT 1
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 12381)
Kellenberger, J.L., Leadlay, P.F., Staunton, J., McArthur, H.A. and
Stutzman-Engwall, K.O.
Polyketides, their preparation, and materials for use therein

JOURNAL

Patent: WO 0001827-A 1 13-JAN-2000;
 KEILNBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
 STANTON JAMES (GB); BIOTICA TECH LTD (GB); PEIER (US); MCARTHUR
 HAMISH ALASTAIR IRRIN (US); STUTZMAN ENGMALD KIM JONELLE (US)

FEATURES

source

1.12381

/organism="Streptomyces avermitilis"

/db xref="taxon:33903"

BASE COUNT 1884 a 4561 c 4005 g 1931 t

Query Match 100.0%; Score 921; DB 6; Length 12381;

Best Local Similarity 100.0%; Pred. No. 1.7e-101;

Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTCTTGTGTTTCCCGACAGGTTGGCATGGCCGGATGGTGTGCTGTGCTGCC 60
 3965 GTCTTGTGTTTCCCGACAGGTTGGCATGGCCGGATGGTGTGCTGTGCTGCC 4024
 61 TCCTCTCCGGTGTTCGCGCGGATGACAGCGTGCAGAGGCTCTGGCCGCGTGG 120
 4025 TCCTCTCCGGTGTTCGCGCGGATGACAGCGTGCAGAGGCTCTGGCCGCGTGG 4084
 121 GACTGTCTGTGTGATCATCTGCGCGGACCGCGGAGTCCGTTGTGAGACGCGCC 180
 4085 GACTGTCTGTGTGATCATCTGCGCGGACCGCGGAGTCCGTTGTGAGACGCGCC 4144
 181 GATGTGTCTGACCTGTGCTTTCAGCGTATAGTGTCTTGTGCTGTGCTGTGCTTC 240
 4145 GATGTGTCTGACCTGTGCTTTCAGCGTATAGTGTCTTGTGCTGTGCTGTGCTTC 4204
 241 TAGCGTATCGAACCCGACGCGGTCTCTTGGCCATTCGAGGCGGAGATCGCGCGCAT 300
 4205 TAGCGTATCGAACCCGACGCGGTCTCTTGGCCATTCGAGGCGGAGATCGCGCGCAT 4264
 301 GTGTGTGGGCGCTGAGCCTGAAAGACGCGCGAAGACTGTTCGCTGCGACGCGCGC 360
 4265 GTGTGTGGGCGCTGAGCCTGAAAGACGCGCGAAGACTGTTCGCTGCGACGCGCGC 4324
 361 CTGGCGGCTGTGGCGGCGCGGCGGATGAGCTTCAATGCGCTGCTGCTGCGAGAGTG 420
 4325 CTGGCGGCTGTGGCGGCGCGGCGGATGAGCTTCAATGCGCTGCTGCTGCGAGAGTG 4384
 421 GACGAGCTATGTGTGAGCGGTGAGCGGCGGCTTGGTGGTGGCGGCTCAACGCGCC 480
 4385 GACGAGCTATGTGTGAGCGGTGAGCGGCGGCTTGGTGGTGGCGGCTCAACGCGCC 4444
 481 CGCTCCACCGCGCTTCTGCGGAGATGCGAGCGGTGAGCGAGGTGCTGCGCTGCTG 540
 4445 CGCTCCACCGCGCTTCTGCGGAGATGCGAGCGGTGAGCGAGGTGCTGCGCTGCTG 4504
 541 GGCACCGGCGGTGCGGCGCGGCGGATGCGGCTGAGCTATGCTGCGACTGCCCATATG 600
 4505 GGCACCGGCGGTGCGGCGCGGCGGATGCGGCTGAGCTATGCTGCGACTGCCCATATG 4564
 601 CAGCCCTGTGCGGAGAGTGTGTGAGCTGTGAGGAGCATGACCGCGAGCGGCTCGCG 660
 4565 CAGCCCTGTGCGGAGAGTGTGTGAGCTGTGAGGAGCATGACCGCGAGCGGCTCGCG 4624
 661 GTGCCGTTCTTCTCAACGATGAGGAGCACTGTGAGCAACAACCTTGAACGCGCGC 720
 4625 GTGCCGTTCTTCTCAACGATGAGGAGCACTGTGAGCAACAACCTTGAACGCGCGC 4684
 721 TACTGTATCGGAACTGTGACAGCGGTGCTGTTTACGAGATGCGTTCAGGCGCTG 780
 4685 TACTGTATCGGAACTGTGACAGCGGTGCTGTTTACGAGATGCGTTCAGGCGCTG 4744
 781 GATGAGGAGACACCGCGCTTGTGAGTACGCGCCCAACCCCTGCGTCCCGCGCATC 840
 4745 GATGAGGAGACACCGCGCTTGTGAGTACGCGCCCAACCCCTGCGTCCCGCGCATC 4804
 841 GAAGACACCAACCGAAGACACCGCGAAGACGTACACCGCATCGGACCTCGCGCGCG 900

Db 4805 GAAGACACCAACCGAAGACACCGCGAAGACGTACACCGCATCGGACCTCGCGCGCG 4864
 QY 901 GACACGACACCGCGCGCTTC 921
 Db 4865 GACACGACACCGCGCGCTTC 4885

RESULT 2
 E38020 30690 bp DNA linear PART 31-JAN-2002

LOCUS E38020 30690 bp DNA linear PART 31-JAN-2002

DEFINITION Avermectin aglycon synthase gene.

ACCESSION E38020.1 GI:18626909

VERSION JP 2000245457-A/1.

KEYWORDS Streptomyces avermitilis.

SOURCE Streptomyces avermitilis

ORGANISM Streptomyces avermitilis

REFERENCE 1 (bases 1 to 30690)

Avermectin aglycon synthase gene

Patent: JP 2000245457-A 1 12-SEP-2000;

THE KITASATO INSTITUTE

OS Streptomyces avermitilis

PN JP 2000245457-A/1

PD 12-SEP-2000

PF 24-FEB-1999 JP 1999046961

PR SATOSHI OMURA, HARUO IKEDA

PI C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC

C1201/68//C07493/22,

PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC

FT Key Location/Qualifiers

FT CDS (1)-(11916)

FT CDS (11971)-(30687).

FEATURES

source 1.30690

/organism="Streptomyces avermitilis"

/db xref="taxon:33903"

BASE COUNT 5355 a 12455 c 8617 g 4263 t

ORIGIN

Query Match 100.0%; Score 921; DB 6; Length 30690;

Best Local Similarity 100.0%; Pred. No. 1.4e-101;

Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTCTTGTGTTTCCCGACAGGTTGGCATGGCCGGATGGTGTGCTGTGCTGCC 60
 3148 GTCTTGTGTTTCCCGACAGGTTGGCATGGCCGGATGGTGTGCTGTGCTGCC 3207
 61 TCCTCTCCGGTGTTCGCGCGGATGACAGCGTGCAGAGGCTCTGGCCGCGTGG 120
 3208 TCCTCTCCGGTGTTCGCGCGGATGACAGCGTGCAGAGGCTCTGGCCGCGTGG 3267
 121 GACTGTCTGTGTGATCATCTGCGCGGACCGCGGAGTCCGTTGTGAGACGCGGC 180
 3268 GACTGTCTGTGTGATCATCTGCGCGGACCGCGGAGTCCGTTGTGAGACGCGGC 3327
 181 GATGTGTCTGACCTGTGCTTTCAGCGTATAGTGTCTTGTGCTGTGCTGTGCTTC 240
 3328 GATGTGTCTGACCTGTGCTTTCAGCGTATAGTGTCTTGTGCTGTGCTGTGCTTC 3387
 241 TAGCGTATCGAAGCCGACGCGGTCTTGGCCATTCGAGGCGGAGATGCGCGCGCAT 300
 3388 TAGCGTATCGAAGCCGACGCGGTCTTGGCCATTCGAGGCGGAGATGCGCGCGCAT 3447
 301 GTGTGTGGGCGCTGAGCTTGAAGACGCGCGAAGACTGTTGCGTGCAGCGCGCG 360
 3448 GTGTGTGGGCGCTGAGCTTGAAGACGCGCGAAGACTGTTGCGTGCAGCGCGCG 3507
 361 CTGGCGGCTGTGCGGCGCGGCGGATGAGCTTCAATGCGCTGCTGCTGCTGCGCA 420
 3508 CTGGCGGCTGTGCGGCGCGGCGGATGAGCTTCAATGCGCTGCTGCTGCTGCGCA 3567

[illegible]

/gene="aveA1"
 /note="contains a loading module, Ave module 1 and module
 2
 multifunctional polyketide synthase"
 /codon_start=1
 /cramel_table=11
 /product="cype 1 polyketide synthase AVES 1"
 /protein_id="BA086474.1"
 /db_xref="GI:5902891"
 /translation="WORDWGGSEPPRAAGEVLGVADADGGVFPVEQGGPQNGR
 BLDSDYDFRESEVCEAFAPVAVMVSVOVLRSDSDAGVLRVVOVPLTAPVMSL
 LAALMSQGVPEPACVAGHSIGETIAAHSVSGSLADAPRVLTMSAOQTLAGVGLS
 VAATPEDELLPRTAPMTDENPAPLAIVAVRPGSTVSGARBAVDIVDLTAOVTR
 MIPVVRPAHSPPLMAVIEBVVSGCLPVTVPERSIPHSSVTSGRDLTDEDAVYTR
 MSSTVRPAHLLIQGGKTVEMSPHVLVTMGDLVDLDDTGTATVVTGTLR
 RGQQLDHPITSLAOLRGGHESATTVLSARLTASPTQOOSLLDLVRANHTAVND
 DNGEATSDAGSAGSAHPIGFPVSGVAGCMGVTALSTKATGLPTLLFDHTPAVAVR
 LRTALGIDETAPVDPSSDSHGGLAAMDPLAIGMCRFGGSRPDMLELAS
 GSDALGEPTRPGTCWTPRBNADHPQPGFPGGGGLHDAAPFADGPGFISREBLA
 MDPOORLLSTSWAPERRAGIDPLSYGSGRTVPAGALSFTDGRNDTASBEADVE
 GHILTGTVSVLSIGIASFGLEGRITVDTCGSALVTLLHACLSGDSCTYALAG
 GVSVMSTGMPTEFISFORASVDGRCASAAADGSGMEGVMLVERLSIDRVLRH
 RVLAVRASAAMVQDASQNLTPNPGAPGRTIROLANALGVADVAVVSGHGTGLT
 GPRTAOLALAYGQAGRPRLPMSISKNGHTMAAGVGIKVMALREVLPET
 LHVDSRPOVMVSACAVLTETAVPMQDARLRAGVSGTGVNHTVLEBA
 AGCCVAGSGVLEAGFGLSVASEBAPVAPVASEVPVPGVTPVPVARSERAG
 LRAQEAALROYAVAPVDSLADVAGALAGRAVLBEHRAVVLADREBLVGLSALAG
 BBRDVTGTHAGGGRGVGVFPVPGQGGCMGVTALSTKATGLPTLLFDHTPAVAVR
 VDMVSVDLIRRAGQADWEBAVVOVPLPESVMSLALRSYSIEDAVVSGHGETLA
 AAVGSGSLAQDAQVATVLAAPVAGRAVRRGMAVSPVLAQVEVQLIGERMAKLWA
 AVNGSRSTVSGSDAANDVYLAKCATGYRARIPIVDASCHENQVPRLEBELD
 LSPQSGVPFESTVGTMVLDITTLDAVATRYLKHQVRSDAQAOLDSDHRAVEVS
 PHPTVVPRIEDTDETPAVALIGSLRRDNDTRRLTALAHHTTGIETPTVHNHY
 THHNHPHHTLIDTPYFQCHVAGHSIGETIAAHSVSGSLADAPRVLTMSAOQT
 EVESFEMVAVADODELTATVAPVPSAGLDTVPALSAHNRHQDOARINTVTOGT
 MKPLTPTTHOQHWMLIAVEPTHNPATITNLTNHHGTRITVLTNHTNPOH
 LHTHTAHTROQONHTGATISYLSILADETRPHNHTPTSTLNTLTQTNTOH
 PPTPLVMAVNTAHTNDRPLTPHQAQWGLAKTLLTSHHTHTGATIDLPPTHTPT
 LOHLTLOTLQPHQQLAIRTGTHTHRTPLTLPTRPHTPHTAGITLIGETGL
 ATHLNHTHTHOPRQTLHTSGSGHTPHACHLTQLOKGLHITTTSCDTSMDQLO
 OLANTIPROHPTHTIHTAGLIDATHTLPTLOLNVKAKHSAHLIOLPHNP
 TAPVLYSSALAPFGAPGVANVAATDALAHNHTHLPATSIAMTQSGSLDS
 DKARVLDREGRRSPSLATVATQALIDTRPVLVADIDMSKLEHRSQISDLVSA
 AKERPAVQRPTRPAREHHTLTAQVSDORALLEVRHVAVLAHRAKQISALAPQS
 FRALDGPETVLEPNLLIKATGLPVSIVDHPRAKLAVILQOLGRTAESLSP
 AAAYVAASVTERIAIVGMACPGPGVTSADDFMPLISEODIIGEPFTGDEIDPL
 YDPRDHGTCTKNGSGLYDAGHDAEFGISPEBALAMDPOORLLSTSWAP
 GINTPTLGTPTGVTCGTNGOYALRVNHAQSTDFRALGTGASVTSIGRTSTPGE
 GRAVSDTACSSILVALHIALCOLARGESVMAAGGVTVMSPGAVERSRQGLAD
 GHCKAFSAADDTGQEGEGVILVRLSPAHNNGHVLAVVRSANODASNGLTAP
 NGPSOORYIROLNAAGSAGVDAVDEAGVETSTTGLPLEAQALATYQDADRAGE
 WLGVSXSGNHTQOAAAGVGVKVMALHGLHLPRLTAVDEBSPHDMWSGAOVLTE
 VLVGSGSGRLRBAARBAEVSCTVKNMALTLEAPADVGGPGRDMSGDEEALAG
 BEGVMPVLYSAKSQALRAQDALAHLTDRGDLADVGTYLAHAPVDRHATLIA
 LDRTDFDLOALAAAGEPRAPVATISSAPGCTGEGAASTAPCSQOQVTRPMAHGL
 YTHHVPFAALNDICTHLDPHDLHPITLPTONDNDENRALLOOTRAQAPLAFQ
 VALHNLIDGNTVTHYVAGHSIGETIAAHSVSGSLADAPRVLTMSAOQTLAGVGL
 MTTHTTTPHHTHTNLTANENADIALINTPSLVSGTHVOHITTLCOQOSIKTPT
 LPTNLAHSPHPNPLINLOHNOTGLTYNHPHTPLTATPREDLLTPHYTQOANT
 VYATATTOGLTHGVTITLEDPTNLTLLTTPHNRPPTTTLTTPHNPHTGLD
 NLAKTTPHNPHTYHNHONQPTHTNHLDTLPYFQCHVAGHSIGETIAAHSVSGSL
 TEHLPLGATLELATDGSGLAGRLSIRSHYMLADAGVTLTGLPTELAHLADAGY
 VGCSDVDELTETLPLVPRDGVSGVOUYAALADBEGRVLSVYASGSGACGGASGG
 WITGASGVVLEAAGAVVVDLACVMPRGVAVDVQVDRBLAAGVCLGAPVGL
 RAVMDGSDDLAECVLPREAMQDAGFGRALDVOVORVSLTPEGCTGFGGAGFEG
 EGVRPEAVAGVSLHRAQTVGRVRSAGRGSGREAVVVGDEBEGVPASVDRLTE
 PRVDQGLAPVAVSAGSRSLTAVUWMAEAVGVPECEQAAMENDLVVERGAVVAGE
 LPRDAGAGGGGGGGGGGGVAGVAGVGVVSGTGLBEAGSRLVVTGRTGAVVAGE
 DGRVAVVAGVWGLVRSADALHNRDEVDLDDTSDTDTGASGAGVGGVAVVAV
 AGCEQLAVASGRLAARLRKESSDVVDVSGDPRASRDVAPORSGGVPARVAVD
 VASGRVLPWLSGGSVLYNGSGVLAALVAAVRLAGVSGDILVNSGRDPAEAGELR
 GELVAAABEAVRIASCDVGRRRVVRLEBVRPAGCTGVNHAHGVLDATATVASTPR
 GETVPAVAAVDAALLDELTRGMBLSFVLEFSSAAGITLGSAGQNVAAANAADALAVR

QY 721 TACTGTACCGCAACCTGCACGACCGGTCCTTTAGCGATGCGTCCAGGCTTGCGC 780
Db 3968 TACTGTACCGCAACCTGCACGACCGGTCCTTTAGCGATGCGTCCAGGCTTGCGC 4027
QY 781 GATGACGAGACACCGGCTTCTGTCGAGTACAGCCCCACCCACCTCTGTCCTCCGATC 840
Db 4028 GATGACGAGACACCGGCTTCTGTCGAGTACAGCCCCACCCACCTCTGTCCTCCGATC 4087
QY 841 GAAGACACGACGAGACACCGGCGAAGACGTACCGCGATCGGACGCTCGCGCGGCG 900
Db 4088 GAAGACACGACGAGACACCGGCGAAGACGTACCGCGATCGGACGCTCGCGCGGCG 4147
QY 901 GACACGACACCGCGCGCTTC 921
Db 4148 GACACGACACCGCGCGCTTC 4168

RESULT 4
E38021 31422 bp DNA linear PAT 31-JUN-2002
LOCUS Avermectin aglycon synthase gene.
ACCESSION E38021.1 GI:18626910
VERSION E38021.1 GI:18626910
KEYWORDS JP 200245457-A/2.
SOURCE Streptomyces avermectilis.
ORGANISM Streptomyces avermectilis.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS Omura, S. and Ikeda, H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 200245457-A 2 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermectilis
PN JP 200245457-A/2
PD 12-SEP-2000
PR 24-FEB-1999 JP 1999046961
PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
C12Q1/68//C07D493/22,
PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1), (14643)
FT CDS (14824), (31419).
FEATURES
Source Location/Qualifiers
1..31422
/organism="Streptomyces avermectilis"
/db_xref="taxon:33903"
BASE COUNT 4136 a 10237 c 11677 g 5372 t
ORIGIN

Query Match 99.8%; Score 919.4; DB 6; Length 31422;
Best Local Similarity 99.9%; Pred. No. 2,1e-101;
Matches 920; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCTTGTGTTTCCCGACAGGCTGGGCACTGGGCGGAGTGGCTGTGCTGCTGCC 60
Db 16630 GTCTTGTGTTTCCCGACAGGCTGGGCACTGGGCGGAGTGGCTGTGCTGCTGCC 16689
QY 61 TCCTCTCCGAGTTCGCGCGGAGATGACGAGTGGAGAGGCTCTGCGGCGCGTGGG 120
Db 16690 TCCTCTCCGAGTTCGCGCGGAGATGACGAGTGGAGAGGCTCTGCGGCGCGTGGG 16749
QY 121 GACTGTCTGTGTGATCATCTGCGCGGAGACGCGGCGGATCGGTGTGGAGCGGCGC 180
Db 16750 GACTGTCTGTGTGATCATCTGCGCGGAGACGCGGCGGATCGGTGTGGAGCGGCGC 16809
QY 181 GATGTGCTCAAGCTGTGCTGTTCAGAGCTATGATGCTCTTGGCTGTCTGTGGCTTC 240
Db 16810 GATGTGCTCAAGCTGTGCTGTTCAGAGCTATGATGCTCTTGGCTGTCTGTGGCTTC 16869

QY 241 TACGATATGAACCCGACCGGTCCTTGGCATATCCAGGCGAGATCGCGGCGCAT 300
Db 16870 TACGATATGAACCCGACCGGTCCTTGGCATATCCAGGCGAGATCGCGGCGCAT 16929
QY 301 GTGTGTGGGCGCTGAGCTTAAGACGCGGCGAAGACTGTGGCTGTGGCAGCGGCGG 360
Db 16930 GTGTGTGGGCGCTGAGCTTAAGACGCGGCGAAGACTGTGGCTGTGGCAGCGGCGG 16989
QY 361 CTGGCGGTGTGCGGCGCGCGGCGGATGAGCTGAGTGCCTGCGCTGCCAGAGAGTG 420
Db 16990 CTGGCGGTGTGCGGCGCGCGGCGGATGAGCTGAGTGCCTGCGCTGCCAGAGAGTG 17049
QY 421 GAGACCTCATTTGTAGAGCGGTGCGCGCGGCTGTGGGTGCGCGGCTCAACGCGCCC 480
Db 17050 GAGACCTCATTTGTAGAGCGGTGCGCGCGGCTGTGGGTGCGCGGCTCAACGCGCCC 17109
QY 481 CGCTCAACCGCGCTCTCGGCGGATGCGGAGCGGTGAGACGAGGTGCTGCGTATCTGTCC 540
Db 17110 CGCTCAACCGCGCTCTCGGCGGATGCGGAGCGGTGAGACGAGGTGCTGCGTATCTGTCC 17169
QY 541 GGACCGGCGGTGCGGCGCGCGGCGGATCCCGGTGAGCTATGCTGCACTGCGCCCATGTC 600
Db 17170 GGACCGGCGGTGCGGCGCGCGGCGGATCCCGGTGAGCTATGCTGCACTGCGCCCATGTC 17229
QY 601 CAGCCCTGCGGAGAGATTTGCTGAGAGTGTGCGGAGACATCAGCCCGAGCGCTCGCGC 660
Db 17230 CAGCCCTGCGGAGAGATTTGCTGAGAGTGTGCGGAGACATCAGCCCGAGCGCTCGCGC 17289
QY 661 GTGCGGTTTCTTCCACGAGTGAAGGACCTGTGAGACCAACCTTGAACGCGCGC 720
Db 17290 GTGCGGTTTCTTCCACGAGTGAAGGACCTGTGAGACCAACCTTGAACGCGCGC 17349
QY 721 TACTGTACCGCAACCTGCACGACCGGTCCTTTAGCGATGCGTCCAGGCTTGCGC 780
Db 17350 TACTGTACCGCAACCTGCACGACCGGTCCTTTAGCGATGCGTCCAGGCTTGCGC 17409
QY 781 GATGACGAGACACCGGCTTCTGTCGAGTACAGCCCCACCCACCTCTGCTCCGCGATC 840
Db 17410 GATGACGAGACACCGGCTTCTGTCGAGTACAGCCCCACCCACCTCTGCTCCGCGATC 17469
QY 841 GAAGACACGACGAGACACCGCGGAGACGTACCGCGATGCGAGCTCGCGCGCGC 900
Db 17470 GAAGACACGACGAGACACCGCGGAGACGTACCGCGATGCGAGCTCGCGCGCGC 17529
QY 901 GACACGACACCGCGCGCTTC 921
Db 17530 GACACGACACCGCGCGCTTC 17550

RESULT 5
AB032367/c 64957 bp DNA linear BCT 18-SEP-1999
LOCUS Streptomyces avermectilis polyketide synthase gene cluster (aveA),
DEFINITION aveA2, aveA3, aveA4 and aveC genes, complete cds.
AB032367
VERSION AB032367.1 GI:5902890
KEYWORDS AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
4; type I polyketide synthase AVES 3; type I polyketide synthase
Streptomyces avermectilis DNA.
SOURCE Streptomyces avermectilis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Organization of the biosynthetic gene cluster for the polyketide
antibiotic macrolide avermectin in Streptomyces avermectilis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
MEDLINE 99380548
REFERENCE 2 (bases 1 to 64957)
AUTHORS Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. and Omura, S.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical

PSHGKGTAAADPLAIIMACRPFQVGSZPOLMEALAGGDAIGPFTDGMGTPEAT
HAODTPOBGFYPOGGGFLHDAHFDGAFPIPSREALANDPOORILLIETMSWAFERA
GIDILSYRSGSTQVPAKALSTFDYGRMDTASBEGADVEGHITLITGTSYLSGKSAYS
FGBEGJLITDTGCSASLVTLHAGVLSQSRSECTFALAGVSDVHPAOFERYIROL
SVDGCKAVSAADGTWGEIIGLAVVSGASVNDGASNGLVAPNBPADERYIROL
ANAGLSADVDDVWVEGHGTGTLGDPTEAKOLATYQOGAGRPMILSLSKNIGHTNA
AAGVGSYKMTMLREGEVLPRTLIVDESPGGLAAGVRLTEFVEWPGDAGELPRAA
GVSSGIGCTANVYILBEAPAGGCVACAGRELEAGPLAISVAVSAPVAPVAPV
SVPVPPVPPVAVPVSAREAGLRAOAKRLROYEAVPOVSLVDAGAGLAROAVLEHH
VYLLACTIONSRSARARTARSSTRAPOARPAPTISARIGLMASSSALITSHANG
GDGGEVAVFVWOGQMAKMGVRLCILTRIVARKMOCEBALVAVWMSVVDILIRDA
GDAVEQDADVQVLPFSVMVSLALMRSYGLIEPNEVJGSHDELAANIYAGLSKDA
AKTAVLPOEVEOLIGERGRLVWVAWGPSTASVSGAEVAVLVCYAGTGRPRTI
PVDASHCPHYOPLRETELLELGDISPOPYVPEPSTVSGWMLTTLTLDAAVYRNHL
OPEVSHVDALDOGHRVLEVSHPRLVADITDTPETRRKOROPREPREH
LLPORLDPYVRRROTHTVAPLIDPRPHLEPPIDAPRPHLPOHONYEESQPRAP
GSGGAGAGSAGSGRATAGTAEBSEREDAVARODLEVATTLVVPBSAGLDIV
PALSMARHODQARINTWYQETWKLPTLTPOQHOQHTWILAEYUHLPHITL
TNLHHGIGTPIPLPLTHHTNPOHLLHHTROHOLMTGATLGLSLADETHPT
PHHPRTGTLNLTLTQHTQTPREPLVATATNAITPHENDELTPHOAQWTLGAR
TLLHPHTAGIILIDPTPTPHLHLTOLLTOLPHOHTOLAINTGHTPHOPLPTL
TTPHOPRTPRHGTITLNGGALATHLTHLTTHOPHLLTSRGTSPHAKOHLT
TLOQXGHLITTTDITSTRPTPLANSANTIPEOPHTVHTGSGIIDDATNLNPT
QLNNVLRKAKASHLHLQTLQHTPLAVLVSAAATFGAOGVNYAANAUYALDALAH
HRHTHLPATSIANGTONGAGNDKRAKALDRGSPMSBELATVICAJLADTER
PYUVALIDMSKIEHTSOTSLVSAAREBAVORPFRPALHTLALHOQADRAL
LEIVRDHYAAVLVRLADPKAIPADQSPRALAGSDILVAFRNLIKATGLRPSVLE
HPTPAKLAHLONLROGRLAESAPSAANVADSVTRPIAVGMAOCPRGVYSADRP
WDLISSBEDAIGCFPTDRGMDLDTLYDBDPHPTCTYTRNGCFYDGHDPDAEFGGIS
PREBALMDPOORLLLETAMETIIEAGINPNHLHGPTGVFTGNGDUYALRVHAGOS
TDEPALTDAGASVIGRISTYFFEGEGPNAVSDTACSSVLVHIALCOLARGESMNL
AGCVYMSPGAIVFESKORGLIADNGKCSAANDGNGEGEGLMILVERLSALTON
GHRVLAVVRGSAVNDGASNGLTADNGSQQORVROLANALNGAGVDAVEAGTGT
TLGDEIAKALITATYGOBRAGEPLMGVSGVNGHQAAGVAVYKMWALRHGILL
PRTLIVBESPHNDMSAGAVOLLTETVMPGCBGRLPRAAGSVSGVSTYANHYLBEA
PADVPGSPRAGEBDSDBEALAGSPCEWMLVSAKOPMLARQOALHMLHDHFG
LILDADVGTILAAHAAVEFHRVYLLAABDTELVQALQALAAEBSRPVYHSAPGCTGT
GEAGKTAFTICSGOQTPRGMAHGLYHHPFEALANDCHLDPHLDHPLTPTOT
LUNDIEDAALLQOTRYAOPALFAQVALLHRTPLHITPYVYNGSLGETGNHLAG
ILTLTDAATLLTORALTMWMPGMLTTLHRTPLHITPLHFNHNDILATAINPSTS
VISGPHVHOHITLCOOQGIKTUUKITPLNNAFHSHPNMLNOLHONLTLYRPH
PLIRANTPDDOLLPHNTYQOARNTVDATTTQHLHOGVNTIIELEGRMDLTLTNH
NLPTANTYTLTLPHHNPHOHLNLNLAKTATTTNPHNYHTHNAOPTEPHMDLPTVP
FOHHNYLLESTOPAGNVSAAGLDPTEDHPLIGATLELITDQGLALRSHPHML
DHAQVYVLGSGATFELALAGTYVEDRDELTELAPLIVPDRGVSUYOVGAAD
GKGRILVSVRARGSGACGGGAGSGWMTCHASGLVLEAAGGVVDGAGVPRGAV
AUNDVDRRLAGAGCVLGPVSGRLPAMVPGDGLAEVCLPEAMAGDAGGLHPL
LNGVYOPSLYLLPGSTGKRGAKGCKGVRYDAVWGYSILRACTYTRVYKVASVNRGG
GREAHSVVDGEGPVASVDRLERLPEVDKAGGAGVSAKORSLSLVAOMAEQVGP
VCGGAMAMHEDVGGGPGVGVVLLRPDQAGAGGAGGAGGAGGVAGVGLAVGOG
WLGERFAGSGRLVVVTRGAVAGPBGDGVVAGVLSWGLYSAADHEPDEFVLLDPT
DTGDLDTGAGAGVVDGGRVAAVVAACEPOLAVRGEALPRLTRESSDDYDAGS
GDTBARSDVPAORSAGVPARSDVSGREVLPMLSGGSVYVGTGRTILGAAYRHIA
GVCCVRDILLYVSRBGPARPAREGLRSPRRTYVACVYGEREYVRLLAGGSCSV
SVFDPBSAVASVTRLPVADSVRKEAADAIAVFESEFYDLYDBDP

Query Match	73.5%	Score 677.2	DB 1	Length 11096
Best Local Similarity	87.9%	Pred. No. 2.1e-72		
Matches	810	Conservative	0	Mismatches 46; Indels 64; Gaps 4;
QY	1	GTCTTCGTGTTCCCGACAGGGTGGCAGTGGGCCGGATGGTGGTCGTCTGC	60	
Db	3328	GTCTTCGTGTTCCCGACAGGGTGGCAGTGGGCCGGATGGTGGTCGTCTGC	3387	
QY	61	TCCTCTCCGGTGTTCGCCGGGGATGCAGCGCTTCGGCCCGTGGGTG	120	
Db	3388	CTCTCCCGGGTGTTCGCCGGGGATGCAGCGCTTCGGCCCGTGGGTG	3447	
QY	121	GACGTGTCTGGTGGACATCTGGCCGGGAGCGGGGAGTGGGTGGGAACGGGCC	180	
Db	3448	GACGTGTCTGGTGGACATCTGGCCGGGAGCGGGGAGTGGGTGGGAACGGGCC	3507	

QY	181	GATGTGGTCCAGCGCTGCTGTTCAGAGGTATAGTGTGCTTGTGCTGTGCTGTGCTGCTTC	240
Db	3508	GATGTAGTCCAGCTGTGACTGTCTTCAAGGTCAATGATGTCTTTTAACTGCTCTTTGGCGTTCC	3567
QY	241	TACGGTATCGAAACCCGACCGCGGTCTTTGGCCATTTCCAGAGCGAGATTCGCGCGCGCAT	300
Db	3568	TACGGTATCGAAACCCCAACGAGGTCTTTGGCCATTTCCAGAGCGAGATTCGCGCGCGCAT	3627
QY	301	GTGTGTGGGGGCGGTGACCTGTGAAGACGCGCGGAAAGCTGTGTGGCTGTGGAGCGGGCG	360
Db	3628	ATATATGGGGGCGGTGACCTGTGAAGACGCGCGGAAAGCTGTG-----3670	
QY	361	CTGGCCCGCTGTGCGGGGCGGGGCGAGCATGTGCTCAAGTGCCTGTGCCCAAGAGGTG	420
Db	3671	-----CGCTCCTCCCAAGAGGTG 3690	
QY	421	GAGCAGCTCATTTGGTGTAGCGGTGGCGGGCGGTGTTGGTGGTGGCGCGGTCAACGCGCC	480
Db	3691	GAGCAGCTCATTTGGTGTAGC---GTGGCGGGCGGTGTTGGTGGTGGCGCGGTCAACGCGCC	3747
QY	481	CGCTCCACCGCGGTCTGGGGGGATGCCAGCGCGGTGTGACAGAGGTGCTGGGTATCTGGCC	540
Db	3748	CGCTCCACCGCGGTCTGGGGGGATGCCAGCGCGGTGTGACAGAGGTGCTGGGTATCTGGCC	3807
QY	541	GGCACCGGGGTGCGGGCGCGCGCGGATCCCGGTGCACTATGCTGTGCACTGCCCCCATGTG	600
Db	3808	GGCACCGGGGTGCGGGCGCGCGCGGATCCCGGTGCACTATGCTGTGCACTGCCCCCATGTG	3864
QY	601	CAGCCCTGTGGGGAGAGTGTGTGAGCTGTGTGGGGGACATCAGCCCGCGAGCGCTCGGC	660
Db	3865	CAGCCCTGTGGGGAGAGTGTGTGAGCTGTGTGGGGGACATCAGCCCGCGAGCGCTCGGC	3924
QY	661	GTGCGCGTCTTCTTCAAGGTGAAGAGGACCTGCGTGTGACACCAACCCCTGGAAGCGCGCC	720
Db	3925	GTGCGCGTCTTCTTCAAGGTGAAGAGGACCTGCGTGTGACACCAACCCCTGGAAGCGCGCC	3984
QY	721	TACTGTATCCGCACACTGTGACACAGCCCGTCCGTTTACAGCATGCGTGTCAAGCCCTGAGC	780
Db	3985	TACTGTATCCGCACACTGTGACACAGCCCGTCCGTTTACAGCATGCGTGTCAAGCCCTGAGC	4044
QY	781	GATGACGGACACCGCGTCTTGTGTGAAGTCAAGCCCCCAACCCACCTGTGTCCCGGCATC	840
Db	4045	GATGACGGACACCGCGTCTTGTGTGAAGTCAAGCCCCCAACCCACCTGTGTCCCGGCATC	4104
QY	841	GAAAGACACACCCGAAAGACACCG--CCGAAGAAGTCAACCGCATTCGAGCGCTCCGCGCGG	899
Db	4105	GAAAGACACACCCGAAAGACACCGCCCGAAGAGCTACTGCAATCGGCGAGCTTCGCGCGCGG	4164
QY	900	CGACCAACGACACCCGCGCGCTTC 921	
Db	4165	CGACCAACGACACCCGCGCTTC 4186	
RESULT 7			
AX211705			
LOCUS	AX211705	65140 bp	DNA
DEFINITION	Sequence 1 from Patent WO0159126.		
ACCESSION	AX211705		
VERSION	AX211705.1		
KEYWORDS	GI:15523937		
ORGANISM	Streptomyces noursei.		
SOURCE	Streptomyces noursei.		
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
AUTHORS	1 (bases 1 to 65140) Zotcheva,S.B., Sekurova,O.N., Fjærsvik,E., Brautaset,T., Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and Gulliksen,O.M.		
TITLE	Gene cluster encoding a nystatin polyketide synthase and its manipulation and utility		
JOURNAL	Patent: WO 0159126-A 16-AUG-2001; Norwegian teknisk Naturvitenskapslige Universitet (NO) ; STIFTELSEN		

IND OG TEKNIISK FORSKNING VED NORGE'S TEKNIISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich
(NO) ; SEKUROVA, Olga Nikolayivna (NO) ; FJÆRVIK, Espen (NO) ;
BRAUTASET, Trygve (NO) ; STROM, Arne Reidar (NO) ; VALLA, Svein
(NO)

FEATURES

source

Location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

BASE COUNT 8270 a 25171 c 22273 g 9426 t

ORIGIN

Query Match 43.7%; Score 402.4; DB 6; Length 65140;
Best Local Similarity 66.2%; Pred. No. 7.8e-40;
Matches 612; Conservative 0; Mismatches 306; Indels 6; Gaps 2;

QY 1 GTCTTGCTGTTTCCCGAGACAGGAGTGGAGCGCGGATGGGTGCTGCTGCC 60
DB 8674 GTCTTGCTGTTTCCCGAGTCCCGAGTGGAGCGCGGATGGGTGCTGCTGCC 8733
QY 61 TCTCTCCGCGTGTCCCGCGGATGGAGCGGTGGAGAGGCTCTGGCCCGGTGG 120
DB 8734 GAGTCGCGCGGTGTCCCGAGCGGATGGAGCGGTGGAGAGGCTCTGGCCCGGTGG 8793
QY 121 GACTGCTGTGTGGAGACATCTGCGCGCGGAGCGCGGAGTGGGTGGAGCGCGGC 180
DB 8794 GACTGCTGTGTGGAGACATCTGCGCGCGGAGCGCGGAGTGGGTGGAGCGCGGC 8853
QY 181 GATGTGCTCAAGCTGTGCTGTTCAGCGTCATGTTCTTTGGCTGCTGTGGCGTTCC 240
DB 8854 GATGTGCTCAAGCTGTGCTGTTCAGCGTCATGTTCTTTGGCTGCTGTGGCGTTCC 8913
QY 241 TACGCTATCGAACCCGACCGCGTCTTGGCCATCCCAAGGCGAGATCCGCGCGCGCAT 300
DB 8914 CGTGTGTGTGTCCGAGTGGAGTGGAGCGGATTCGACAGGTGAGATCGCTGCGCGGTG 8973
QY 301 GTGTGTGGGCGCTGAGACCTGAGAGACGCGCGAGACATGTTGCGTGGCGAGCGCGGCG 360
DB 8974 GTGTGTGGGCGCTGAGACCTGAGAGACGCGCGAGACATGTTGCGTGGCGAGCGCGGCG 9033
QY 361 CTGG---CCGCTGTGGCGGCGCGGCGCGGATGCTCAGTGGCGTCTGCTGCCAGAG 417
DB 9034 ATGTGCTGTGGCGGCGCGGCGCGGATGCTCAGTGGCGTCTGCTGCCAGAG 9093
QY 418 GTGGACAGCTCATTTGTGAGCGGTGGCGCGGCGGTGTGGGTGGCGCGGTCAACGGC 477
DB 9094 CTGGAACCGGCGGTGTGCA--GTTGAGAGGCGGCGGTGTGGGTGGCGCGGTCAACGGC 9150
QY 478 CCCGCTCCACCGCGCTGTGGGAGATCCGAGCGGAGTGGAGAGGTGCTGGCTAGTGTG 537
DB 9151 CGCGCTCCGCTGTGGTGTGGCGGAGCGCGGAGCGGTGAGCGCTGACCGCCGCGCTG 9210
QY 538 GCCGAGACCGGAGTGGCGGCGCGGATCCCGGTGAGTATGCTCTGCACTGCCGCCCAT 597
DB 9211 ACCGCGAGACATCCGCGGCGCGCGGATGCGGTGAGTATGCGCTGCACTGCCAGCAG 9270
QY 598 GTGACAGCCCTGCGGAGAGATTGCTGAGACTCTGCGGAGACATCAGCCCGAGCGCTGC 657
DB 9271 GTGAGAGACCTGAGACAGGAACTGTGAGAGTCTGCGGAGCTGAGCGCGCGCACAGTGTG 9330
QY 658 GGGGTGCGGCTTCTTCTCAAGGTGAGAGGAGCACTGGCTGGAGACCAACCCCTGGAGCGGC 717
DB 9331 GAGGTGCGGCTTCTTCTGAGCGGTGAGCGGAGCTGGCTGGAGACCAACCCCTGGAGCGGC 9390
QY 718 GCGTACTGTGATCCGCACTGACAGCGCGGTCTGTTTCAAGCATGCGGTCCAGCGCGCTG 777
DB 9391 GCGTACTGTGATCCGCACTGAGCGGAGCGGAGTGGGTGCGGAGCGGAGTGGCGAGCTG 9450
QY 778 GCGGATGAGCGAGACCGCGGTCTTGTGAGAGTACGCCGCCACCCACCTCTGTCGCCGCC 837
DB 9451 CTGGCGCGGAGTACCGCGGTCTGTCAGAGTCACTGCAACCCGCTGTCAGAGATGGCG 9510

QY 838 ATCGAAGACACACCGAAGACCGCGCGAAGAGCTCACCGCATCGCAGCTCCGCGC 897
DB 9511 GTCTTGACCTGATGAGAGAGCGCGGAGTCAAGCGCGGTGCGAGACCGGACCTGCGCGCT 9570
QY 898 GCGGACACGACACCGCGCGCTTC 921
DB 9571 GACGAGGTTGCGCGCGCGCTTC 9594

RESULT 8

AP263912

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

FEATURES

source

1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement(46..783)
/gene="nysF"
complement(46..783)
/gene="nysF"
complement(46..783)
/function="presumably post-translationally modifies ACP domains on PKs"
/note="putative 4'-phosphopantetheine transferase"

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

1..123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
complement(46..783)
/gene="nysF"
complement(46..783)
/gene="nysF"
complement(46..783)
/function="presumably post-translationally modifies ACP domains on PKs"
/note="putative 4'-phosphopantetheine transferase"

transl_table=1
product="NysF"
protein_id="AA071762.1"
/db_xref="GI:8050836"
translation="MIELPLPATVTEAAYDDRPRGRLLSSEREVIARAVERSHORE
FTTVHLARLRRLRGHPDRAIIPNRGAPMPGIVGSMTHCAGYRAAASPARLSA
AVSIDAEPNGPAGVNAIIPSRRLVLAARPDVHMDRLLPKESYKAWP
LTORLDSSEABIVDPDGAFTARLIVGPIUGGRVTVFPGRWHSPTALTTVAHL
PAPYRRDRERHRLTVNSPLPRPTG"
complement(867..2684)
/gene="nysG"
complement(867..2684)
/gene="nysG"
complement(867..2684)
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
codon_start=1
transl_table=1
product="NysG"
protein_id="AA071763.1"
/db_xref="GI:8050837"
translation="MASPDLEERTAPARLVGLRPHRSVALAVSMGVGIVL
NAFGLLIGRVTLIDAVLGVPAGPAGIDPAALIGRLVLAALVYVAALFPLAOG
LVAASVMTIHRLRDAREKULRLPLRHFDRQPAQELLSRTNDIDNLOOTLQOTLAE


```

QY 718 GCCACTGATGACCGCACTGCAACAGCCGCTTCCTTTCAGCGATGCCCTCCAGGCCCTTG 777
| | | | |
Db 69652 GGCTACTAGTTCGCCGCAACCTGCGGGAGCGGGTCCGGACCGCGTGGGGAACCTG 69711
| | | | |
QY 778 GCGGATGACGCAACCGCGCTTCCTTTCGTAAGTCAAGCCCAACCCCTTCCTCCGCC 837
| | | | |
Db 69712 CTGGCGCGGAGTACCGCGGCTTCCTGAGTCAAGCTCCGACCCGCTGCTGCAATGCGC 69771
| | | | |
QY 838 ATCGAAGACACCGACCGAAGACACCGCGGAGTCAACCGCGATGCCGAGCTCCGCCGC 897
| | | | |
Db 69772 GTTTCGACCTGATTCAGAGAGCGCGGAGTCAAGCGCTCCGACCGACCTTCGCCGT 69831
| | | | |
QY 898 GCGGACACGCAACCGCGCTTC 921
| | | | |
Db 69832 GACCGAGGTGCGCGCGGCTTC 9855
| | | | |

RESULT 10
AB070940 104326 bp DNA linear BCT 22-JAN-2002
LOCUS Streptomyces avermitilis oligomycin biosynthetic gene cluster.
DEFINITION AB070940
ACCESSION AB070940
VERSION AB070940.1 GI:15823967
KEYWORDS Streptomyces avermitilis DNA.
SOURCE Streptomyces avermitilis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces; Streptomycetaceae; Streptomyces.

REFERENCE
1 Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinoe, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T.,
Kikuchi, H., Shibata, T., Sakaki, Y. and Hattori, M.
Shinoe, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T.,
Kikuchi, H., Shibata, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
2 (bases 1 to 104326)
Ikeda, H.
Direct Submission
Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
Fax: +81-3-3444-6197)

FEATURES
source location/Qualifiers
1. 104326
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
complement(1..1134)
/gene="mcma"
complement(1..1134)
/gene="mcma"
/feature="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=11
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
/translation="MTGYSTADANARFQVSGSTRVAVVFDLETRAGLDSADAPAA
RGLVGRGVAADSIDMRVLFGLRDRVPSLADAPAPLVLVQLVAERHGAPO
RAGAVRGVFEFLARGPVPPLRSLANVAVYCOAELPRMOCLESVGHLAGT
GADPALEVAFTVAGGIEYLAATAGLDRVAVAPRLTCLVGGTGERAGLRATRR
VMRTANERFGSGRTAPRLVYRTPRGLATLTACTBTEIGATVADRERLGGTTLV
EHGLAQLGRPHGRPATRRPHPPQDSLDIARTEERLAKIRANRVDPAYDEALLR
LTKAAGDNDNVLPMEKALAAATVGEVCDALRTWTGTPSAT"
complement(1131..1886)
/gene="pPCA"
complement(1131..1886)
/gene="pPCA"
/codon_start=1
/transl_table=11
/protein_id="BAB69186.1"
/feature="phosphopantetheinyl transferase"
/protein_id="BAB69186.1"

gene
CDS
/db_xref="GI:15823969"
/translation="KRAHVFADVPVSSLHPREALVODAVESRRREFATGRHCARA
IASLVPPPELILATPRGAPRMVAGVCGATTHGAAVAARAPGIRAGIDAEEDRP
LPEGLVNAALPAEOPVYKENTARRAPVWGKLLTSAESVYKAPFATGRPLGEEA
LIDLDVSTFRFVRVPLTALRAABELDLHMRMLVRHGIITAVAVVDGPGDTHAP
SPYDRTKHSAFPGRSADRPGRPTDAPRPA"
complement(1910..2875)
/gene="simx4 homolog"
complement(1910..2875)
/gene="simx4 homolog"
/codon_start=1
/transl_table=11
/product="Simx4 homolog"
/protein_id="BAB69187.1"
/db_xref="GI:15823970"
/translation="MSQRHDSSTAALGSPSHGRLLAIVSDIHVRHNRNDVLDLP
GAEWDLIVAGDVASVADITWALSLELDRAKVIWAGNHELTAPGPDLEBERK
YKHLVDLCRALGVVTPEDPYVWGHAGGVAVAPFLLYDSFLPFGMAKSHALRLA
DDAGVICTDEFLHPDPYPSREAWCARAVATRALDLPDLPTVLVNHPELVPEPT
RVLRRPEALMCGTATADWHLRYRAAIVVGHGLHIPTIRICDDVPVHESLGYPREW
ORYGRPPMPVQIILTPAPAPAPVVPVPAPEGHRTASPADRR"
3213..3971
/gene="OlmC"
3213..3971
/gene="OlmC"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="BAB69188.1"
/db_xref="GI:15823971"
/translation="MSANRVGNTLVROFHPATATKLVCLPFGASASAYFLPYSKA
LAPGDVATVOPGDRRHEKIDNIKELDAVAVLEPLADRLVLFEGSMGTLA
FEVALRLRKGIPLALFASGRBARSHRSDLRGDEIILKTLTSGRSEVLDG
BEILRMVPAIRSDYKAAETRYAGATLTPPIHAFVEEDPRVSLDQAQMAHDITG
HPELRSHPGHFLYMDQAPVIEETARVSSVAS"
4643..7438
/gene="OlmR1"
4643..7438
/gene="OlmR1"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69189.1"
/db_xref="GI:15823972"
/translation="WLTGRGQIOLRLRAVSVSTGRSAVIVSAGCGKSEILD
LVRYVESGLIALRTALAEERNFTLSTRBLVNCAPPPDDLYRRFHTMSDPAAMIS
TEGRDKARDAWELMKRTCTEVVALAQAPLVAVVDIHGDSDLSRLQLHARHC
RNVPVLTFTQSLAGPADRTFGALRLQPNLEHILRLRPFQODIADMLADRGADAG
OGTVRYFAVSGNPLRLALLEDLRLEDDVDVQDVGPRAGAVYGRAVACLDGCG
EKAEHVAAGLAVLGAFTHEILSRLLDISAALTENLALRAAGIVGTSFOHPAKA
AYLSLEPSHRLRLHORSALILHAGAPALAAEHLHRAAGADGAMAVGVIQAAEKA
ILADEARAIVSILELNAASADPRGRTIIRLUSVTRVRSADDERVDDALRFGA
GILTPSDTGLAALLLAHGRLEQASVLEQRRPDDGDADRPAARQLAERDVLNR
GAWCAGTGAAMHNPAGRPSTGAGTGLRAMLGSHALIAQRLAGTEGLLEITLTD
ATPDLVLAIVVLTILAGVERALHWCDFLEAKRRDAPGWEVAVAARAGALRGAN
LABARDARCPAVLPERHRSALGSAVSIVTVQIARKEYEDAVAROLSRPADSLPD
TTHMDVYARGLVYMATROYHSAIDRHYVGRALRERGTPDALLPRTGTALR
IGEROGCRLATBOLAMVQDAPRKISLISLQGLASPARRLILINAVYBELSAD
RLIESRTFLDLADARETRQTOAMVIRARQLAKSGCAOPRRISPPAAGSAGE
LFGEPENNAWMGHLSSEBSERVAGLAAGYTNREISVLYITMSTVEOHLRVYKXK
INGREPLAEARAGQV"
7658..10486
/gene="OlmR11"
7658..10486
/gene="OlmR11"
/codon_start=1
/transl_table=11
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69190.1"
/db_xref="GI:15823973"
/translation="MLASAGQGEQGVIVNAGPVGRTALLDAFLRKATSGARVCA
AGSSAETAEFGVGOLEPPPGPLGAGRLARAGSRAGCGADRLADRLFDLCELR
RPTVIAVDQAQDAASIRFLRYIVRRLRTAPVMVVLTEPGDALPSPFOALLRHP

```


DEFINITION Streptomyces nodosus amphotericin biosynthetic gene cluster,
complete sequence.
ACCESSION AF357202
VERSION AF357202.1 GI:14794889
KEYWORDS
SOURCE Streptomyces nodosus.
ORGANISM Streptomyces nodosus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11393)
Gaffrey, P., Lynch, S.V., Flood, E.M., Finnian, S.M. and O'Leary, M.
The amphotericin biosynthetic gene cluster from Streptomyces
nodosus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 11393)
Gaffrey, P., Lynch, S.V., Flood, E.M., Finnian, S.M. and O'Leary, M.
AUTHORS Direct Submission
TITLE Submitted (07-MAR-2001) Industrial Microbiology, University College
JOURNAL Dublin, Belfast, Ireland
location/Qualifiers
FEATURES
source 1. 113193
/organism="Streptomyces nodosus"
/db_xref="taxon:40318"
/complement(4. .1824)
/gene="amphG"
/complement(4. .1824)
/gene="amphG"
/note="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="AmphG"
/protein_id="AAK73498.1"
/db_xref="GI:14794890"
/translation="MAPDVEHEEERESROPVRLAALIPRRBSVGLATAGVGI
LNAEREPILGRVTDIADGVUGHGPARCVDGALGRLLMILVIVYVAFMYLON
WLVASVRLIHDLRADGKRLPLRFRPAGETLSRGTDVDNIOQTLOQT
DISSVFSVIMLSMLIISPISLAGVMLSVPSGLAAMISKRAQPOVNAQIT
LTAHVEEMGAGALVAFDRPAEAEQFEDRENAVYRAGSGAOFASGALIEPVMMFVAV
LGVAVAVGAMKVNGSLTIGVQAFITVYARQFQPIVEIASVAGRLSGVSAQOV
FTLIDAPROEPDRPLAVERVGRVPEQDVSRYSPDRLTGLSLSTSPSTAVV
GRSGAKTYANILNRPYRIDSGRIILDDGDTTAAWRDIDRSFGVLDOTMFKETI
AENIATGSPGATADIVEARATYADRFRITLSQGYTVLIDSSGVSAGEQITIVA
RMLARPAVVLIDEATSVDTRETLICRAMMTLRGRITSFVIAHRLSTIRADAVIV
MSGRIVEQGTQHDILDAQGAAYARLAARADAPAAADVTG"
/complement(1805. .3628)
/gene="amphH"
/complement(1805. .3628)
/gene="amphH"
/note="possible amphotericin export protein"
/codon_start=1
/transl_table=11
/product="AmphH"
/protein_id="AAK73499.1"
/db_xref="GI:14794891"
/translation="MAPSVRWPTPQRAAAGRRQPCSHVLLMKTOLRPYAGSVVA
LVTLHVOILGTLPLTGAALIDEGVHRDSDRIGTIGATMAVVALVOIAALGAA
LGRSTALGRDLRSAPFRVLDFESARETGRFSTPLTRSVNDVQOVNLOQTGRI
FVAAPIMCISVILAROVDTALILVPMVIVAVGEGILLRMAIVLALVQOTDRI
GRLLRRTIGVRYRSFARDHGERFTTNEELSLIGVGLIIVMLPSVILLNL
FTVGLIHWARRIDSGSQIGALSAFLSTSLILMSVLMVAFVILNPPARCAERT
EVIQAETDVPPASPRPMAGQVELVGAEPFPAENAVLRDLSTLRPBERVAVL
ASNGRGTDLTLHLILRLVDTAGEVRIIGTVDLELDPVMAAVGVPPRPLFAGTV
GSMRGKRDPADEELMEVLRVIAQHEGFPVRLGGLDTEIAOGGTTVSGGQORLIAR
ALLRPAVILFDSPALDOSTEALARKALVPTBGAITYTVAORVASRDADRIYLL
DGGIATATGTHDLALDSTPYREIALSQRTRETAAGRS"
gene 3840. .4874
/gene="amphdIII"
CDS 3840. .4874
/gene="amphdIII"
/note="mycosamine biosynthesis protein"
/codon_start=1
/transl_table=11
/product="AmphdIII"

/protein_id="AAK73500.1"
/db_xref="GI:14794892"
/translation="MPKRALITGITGDDGSYLAEHLISGQVOMGLIRGOANPRKERY
SRLASELSVDDGLMDGOSIVSAVDKQDPDEYVNIKAISFPMMSQOAEITVEWNGG
VIRVILBAITMSGLSMRTAGTEGQIRTPQASSSMFGKVAETPQRETTLPFRPBYG
AKAGAHPTIRNRYRESFGMYAVSGMLFHNESRRQGFETRKISLAIVAKIKGLDCKL
ALGNMDVADWQVAGDVYRANMLMQODAADDYVIGTGMHTVRVAVAFAPFVGLDW
KQYVVDVPLVPAVEVILCADSSKAQQLGWKPSVDPELMRMVADLASVRONE
LDLLLASWS"
gene 5042. .33574
/gene="amphI"
CDS 5042. .33574
/gene="amphI"
/note="polyketide synthase multienzyme polypeptide housing
extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/transl_table=11
/product="AmphI"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="WDNEOKLRDYKLATADLRARRRVELESASOEPIATVGTCTC
YCGVSPEDLMRWYENGCVTPPTPTGNDLELASSPSSRGFLDAPRQDF
GISPREAAMDPOQVLESAMEATERGINTPTSKGRTGVFTGMAQDRVGRADG
AGFQLTGNTGSVLSGRISYTFGVGPAVVDVTAQSSLVAVHATOLRAGECTAL
AGGVTVMSGPGFTFEMGRQGLSDVGRCSFEDTADGTGMEGVGILVIERLSDAIRN
GREILAVRGTVANODGASNGLTAPNGSQAQVBOALVNAISAGDIDVVAHGTV
TLGDPEAQAALATYGOORDEKPKLLLSVSNISHTQAAGVAVITMMAMRHGVL
PRTILADPRTHTVMSQAGVYLTENTMTPGAPRAAVASFGISGTHATVTCAP
EPEDPADPDASTPAVAVGVLPILSGRSEVLRQAQALITLTGTTPAPADLAY
SLATRTIAEHNHAAVLASDPELTGRLTAIAGDTPAVLADVTGTATETRLAVALF
QGOAQRLAGRELARFPVFAALDAALDAFLPHLDVLRKLMGEDRDRDRTYAO
PALFAVEVALYRLSESFEVKPDHLAHSVGEIAAIVAGVSLDAALTVARGLMO
ALPEGAAVAVOASEDEVAAPILAGHEDDILVAGVAVVSGEYVTEALAA
DGRKSRVAVSAPFSPIMAPMLDSERVVSLTLAGSLPLVPSVGTPEVAVDLS
PDYVDHRAQVAFADGIDMLARHDVTFLEIGPVSYSAMAQNLDAAGSDALTVA
LREGEDEHTGALAAALHTQGTALHMCACFTGTRARTDPTPTVFORRWPRVAVG
GAADRSVGLGAALHPLTAAVSLADSGALCTGTGRTDPTPTVFORRWPRVAVG
ADLEIAVAGVGCDDVDELTAAPLVLEPGGVQVOMLGNPASGRSVTVRGLQ
FADDEAPMTHSHTVGLSASRTTSDATVWPPADKELPVGLIYERLAEGGGVGLP
FCGRVAMRGSEVFAEVLPSGSGDASFCGLHPLDLSALHASFVLDDEAAGCLP
PSMEGVSLHSGATTLRVRLAPAGDAVAIAYADDSGLVLSADSILIRAVARIDTA
AAALVRDLFRIDWPVTVAVASGTAALVEBDEPGLALPOGLAVHPDLADAAA
DGAVADTVLLPLTGGPDADPVTAAHRAATRAALVRLRQDERFAASRALVTRGA
TGGHPAPAAVWGLVRSASQENPGRFLVLDADQDTLPAALSTSEPCOLAVGEE
LRAPLVRPASTAEPVAFPGEGEAVLVGTGGGAGVLAHLVHEGVRELIVSR
GGAAGAIEVAIEIAESGAPATVACVTDPSAAVEIAVHAPVSAVHSACTDDGMV
GTITPRLITTVLRPVDAANILHEATRLDILKAPLVBSSVAGVGSPOQAYTAAGNAF
LDALAAHRRAGLPELISLAWGPWEQTGNTGSIISDDIRKRAVAGTALVVOGLALA
DAAIDGDDAALAPVRLDLSVLRAGEVPEPLRLSLGRSRAVAVASATAGLQRLA
RLDASRDELVIDVRGOVALVILGHATGAEIDAGAFRELFDSLTALVELNRNLTVT
GLRLPATLVFDPPTVSHLSYVLDLITGEVBAEVORGTAAVADDPILVIGMACRYP
GCVTSPEDLMLVTBGTDAVSGFPVNRMDVNIYHPDDHRTGTVTSGEITRAGE
FDGPFSGSPREBALTDGQORLLBASWEALBRAGIDVGLRGATGVYRAGVMSDYS
AMLGSPFEFGFQSGSSPSLASGRVSYLIGEGPVAIVDTACSSLVAMHAAQOLRS
GEISLALAGVTVWSTPVAIVDFARORLSTDPGRKASASDADGVMGEGVGLVLER
OSDAILRMHOLIAVTVGSAVNOGASVPLDGPNGSQAQVBOALVNAISAGDIDVVA
HAGGTTPEQIEAQAALATYGRDRDEQPLLLGSVSNISHTQAAGVAVITMMAMRHGVL
SMRHQVLRITLIVARBSHVMTBGAVALVLEQRTAMPETGSPRRVAVSFGISGTH
TVLEBAGCTVPAPAPERTAGVADREGLASGLAARPPVAVGSAARARAVAWFSG
GSGCAAGMRREIYEVORPFAAFDEVCALDPLRPREVFAVAGESEALDRLV
LMEPLFAVEVALYRLSESFEVKPDHLAHSVGEIAAIVAGVSLDAALTVARGLMO
ALDTLRVRENTSRKPAAPVLDGTLVSSDPAIDETELDLSALSGMRQEK
RLVTDACADRAVIGARLADTENADNTAQVSLVLDKRPDGPAGFPOGLATIALV
RGRKTRKLRVSHAFHSPHMDMLAFRVAATLSYAAARITVLSLTERRADDEAVRT
ADYWRVREARFADCVRTIRIDAGATLPLELGGGLITAAEDTLGEBRDHNTALV
PLRADRPEESAAATAARLIQIHGVDDTLVLTAGTGRARVDLPYAOAHAWFOLP
SAAPSPADPADQKLMAAVERGDALEIYALJEDSDITPLDLSLMSWRQEK
ALDITLRVRENTSRKPAAPVLDGTLVSSDPAIDETELDLSALSGMRQEK
RLVTDACADRAVIGARLADTENADNTAQVSLVLDKRPDGPAGFPOGLATIALV
CALADTGAHRLWTRTRGAVSTGRADPPTTHRAQATRWKMGCVLAEHRRNGVLLP
ADPDRGAQORLAIEVLAVDADPDEQVRLATGVGRRLVHIHIDELSAQDTFSGS
VLITGGTGLGAEIARMLARSGAHLVLTSGRDPAPAPAAVRLAELESGASVSLVAC
DVADRDAALAVLDIGLSADQPLTGVTHTAGVGHYGLDALTPAEFAGLTAALGAAHL

DNLLGRELDFLLFGLSAGWGSQDQSVGAANAYLALALARRAGLAATSIAMGP
WGGTGMADDAVSGTLRQGLGLDPAALTEMRRAVIVODVTVADADVYAFLE
TSARPSALISDLEPEBALAENTPADTGAASEIYORRSEIPEBODLJTDVRETA
TVLGSAGAVEPDRAPRRIGPDSLTAVELRKLGAATGSLPSTWVPVYPRLEAO
YLRABWVGSVLEAVGPAVGTGDEPIALIGNSCRIPGVSSPEBQMDVLSTGDTIT
DFPNVGNMTAGLYDPDPHGETTISTQGFLHAEDEPFMPFSGISPRALVNDPQOR
LLEETWEAEBRAGLTPDLRGLSTGTGSSYQXEBNGENLAVAGATMTTSPSLA
GRLAYFGLPRAVIVDTACSSSLVALHLACQALRNGENLAVAGATMTTSPSLA
PSRQALADGRCARKEASADMTLAEQGLVLEVESLARRRQGLVILVIGSALNO
DGSNGLSA PNGSOORVIROLANARVAPGEILDLEAGTGPGLDPIEADLPATY
GRTTRPETALLGSYKSNIGHSQSAQVASTIKMVALHGMPTLHADDESSHYDM
SGETVALGENTDWPQTRPKRAVSSFCISGTNAVILIEQTEAPAEDEBOLAPVL
PVAAGVPMWLSARGAALREBODRLTLVADPAARIDIGLSLSTARRLEHVA
VVPACDTDEALRAVADPGSVARGADVATVFPFGSGQSMAMQMDLDES
PVAERIAACCAALAEFTDMNLIDVLRGAGATLRVVVOVPGSAVMAVLAARA
OGVEPAVVGSGOGEIAAVVGSGLSRGARRVITRAQIRAGSLAGRCQMSVIALPV
AVERBARLEPGRGVVAENGRSSSVVAGPEALDELHQLTNEEIRARVAVDASH
SHVEDLHELELELAVAPRTSEIPFESTVGTGMDLTTVMAGVYRSLRGVLPAD
AVRDLIAADHRAFEVSSHVPVAMSQMDIDAGVAGVAGTLRNDGDLDFLLSA
EVRVGVQVDMAAVEPGETGASVLDLPTVAFQHEENLMAAAEAVATVADPEDAFMTA
VEDQVSALTALGDEDSVAVLPALSSMRARRKRSIVDSMRVPTWKPVTKLPOR
TIDGTMVLVSADGVDVDALELTCGASVRRLVLDSECTDPAVRERLTDDGLTGI
VSVLAGERTGAVPRTGLVGLVALTYVALQALDAGIDTPVLAITRGAVSTGRSRYT
AVQAVGTIGWTALECPGRWGVVDLEPTIDARQGLAVALAGALDQDLARS
SGVETRRIVRADAAPDGSARDKPRGTTLVGTGSGTLADHLARMLAEQAEHLVIVSR
RGEPAERGAELRAELREGTEETLAACDITDDAVALLESIAEERTVTRVHTAT
IEHLTDAITLDDDFRVLAKVTGAQILDELDDELDELPVYSSTRVAGMGGAHAY
VGNAYLALAEHRBARGLTALISWICMAADLOAGRVDPQMRIRBGLFEMPOLALS
GLKRALDDEBOYIAVADVMEYTHPYTARPPULDEYPEROLITVAAEQSGADAR
GEFAALLALPAABOHRKLETVRETAASVILGSSABEDTDORAFDDVDFSLTAVGL
RRRLASVTGLTLPSTWVPYPMPLAGVLSHGLVADHSAVAVATGAPVDDPITAI
VMSCRYPGSITSEQLMRVLEEVDVAFVPADMRDEALAYDPDPDGSQSTYSQI
GRLDVAADPGPGFISPREALSMOPOQLLEETAEVEMNGLDVADRGRTGRTFI
GASDYDGAAPVSGSGSEGHMTIGSLPULSGRVSLRELBEPALVLDTRACSSLYAI
HLACQSLRNGESTLALGASISTMTMSTGSRQALAEDEGRCAIAGAGMTLAE
GAGLILRLSDARRNGEVLAVIRGSAVNOQASGLAPVSGSQRIYRALLANAG
VEANDIVLEGHGTALDPIEALAFATYGDROPEPVLVLSGYSKSNIGTOMANG
VASIILKLVHAREGVAPKSLIDOPSTHMSGTQLTTEREMETGPRPRAVSS
FGLSGNHTVLEQAPADAPAEOTPARDALPVLVSGRGAALRAQAGALLDLIA
ERRGIHPDILASITSRALEHRAVAVDHDHLEVRGTLRDLGPGVIOGRTGR
GRTAPLFTQSGSORLDMKRELYERHVPFADLDVALYRIDGTERTRVLYLPABESQ
DUALHRTYADPALFALVALFRLESVGVTPDYLAGSVEILAAHAGVLDLDDA
CTLVARGRLMQLPEGAMVALLEAEDVLPILBELAQOVVAANNGRSGVVAEVE
BEVLALAAHFAEGCRRTKLRVSAHFSHSLMDPMLDFAVLAALYHAPSLPFSNV
TGTLAPEOVCTADVWVSHVRSVAPRADIIGMLSTGQOTPELGGPDVLGSMABES
LDAERTALLPTLRGDRPEQALVYVAHAAGHPVDVMTAPQSGARVAPLPTAR
OBERVYPTTAAIGTRAPAGSALDAEFNAVEHADVASLJASGLDADATVMPVLT
ARRORGEQSALDSWRYRTWKPRGAPAPAPGRVILVPEHRDEATAAVALDVEA
ALATVAVVEVGTDRALALARLEADADDTGCVLSLALAPGDHGPAPALTLT
ATLQALGDARIDAPLMNITRGAVAGVEOYTAPEOALVMGLFRALAELEPARVGS
VLPEDLTOARIRLGRGILAAADGEBAVAVRASGVRLRLASPADTVGSAFDPAG
TVLITGCTGIGGHLARLAROGAALLTTSKRGDPAAPAGELRAELBSGARVITIA
CMAADRALALALATVPEBDAPLVAVHTAGVDDHVDDELTPESFTVLAHTVAAHR

Query Match 43.5%; Score 401; DB 1; Length 113193;
Best Local Similarity 64.7%; Pred. No. 9.9e-40;
Matches 596; Conservative 0; Mismatches 325; Indels 0; Gaps 0;

QY 1 GTCCTGCTGTTCCCGAGAGGCTGGCAGTGGCCGCGGATGGGTGCTGCTGCC 60
DB 16592 GTCCTGCTGTTCCCGCGGAGGCTGCCAGTGGCCGCGGATGGGTGCTGCGAC 16651
QY 61 TCTCTCGGGTGTGGCCGGGAGATGAGGCGGTGAGAGAGCTGTGGCCCGCTGGGG 120
DB 16652 GACTCCCCGGGTGTGGCCGAGCGGATGCGCGAATGCGCGCCGCTCGCGAGTTCACC 16711
QY 121 GACTGCTGTGTGAGACATCTCTCGCCGCGGAGCGCGGGAGTCCGTGTGGAGCGCGCC 180
DB 16712 GACTGGAACCTATGACGCTCTCGGGGGGCGCGAGGGGCGCCGACGCTGAGCGCGCTC 16771
QY 181 GATGTGCTCAAGCTGTGCTTCAACGCTCATGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
DB 16772 GACGTGCTCAAGCCCTCTGTTGCGCGTGAATGTCTCCCTGGCCGCGCTGTGGCGGGCC 16831

QY 241 TACGGTATGAAACCCGACCGCGTCTCTGGCCATTCCACAGGCGAGATCGCGCGCGCAT 300
DB 16832 CAGGCTGTGAAACCCGACCGCGTGTGTGCGGCACCTCGACAGGCGAGATCCCGCGCGGTG 16891
QY 301 GTGTGTGGGCGCTGAGCCTGAAGAACCGCGGAAAGACTGTTCGTGGCGAGCCGCGCG 360
DB 16892 GTCTCCGCGCGCTGTGCTGCGCGCACCGCGCGCGCGGTGTGTGACGCTCGCGCGCGAGCC 16951
QY 361 CTGGCCCTGTGTCGGGGCGCGCGCGCGCATAGGCTCAGTCCCTGCTCTCCACAGAGGTG 420
DB 16952 ATCGCGCGCGAGCTGTGCGCGCGCGCGCGCATGATGTGCTGTGCGCTGCTGCGCGAG 17011
QY 421 GAGCAGCTCATGTGTGAGCGGCTGTGGCGCGCGGCGGCTGTGGGCGCGGTCAACGCGCCC 480
DB 17012 GTGAGAGCCCGCTGTGAGAGCGCTTCAAGGACCGGCTGTCCGTGCGCGCGAGAACGCGCCC 17071
QY 481 CGCTCCACCGCGCTGTGCGGAGATCCGAGCGGTGAGCAGAGTGTGCGCTGCTGTGCGC 540
DB 17072 CGCTCTCGCTGTGCTGCGGCGCGAGCGCGCGAGCGCTTGAACGACGACAGCTCAC 17131
QY 541 GCGACCGGGGTGCGGGCGCGCGGATCCCGGTGCACTATGCTTCCGACCTGCCCATGTG 600
DB 17132 GCGGAGAGATCCGGGCGCGCGGTGCGCGGTGCACTACGCTCGCATCCCGCATGTG 17191
QY 601 CAGCCCTGCGGAGAGATGTGCTGAGCTGTGGGAGCATCAGCCCGGAGCGTCCGCG 660
DB 17192 GAGGACCTGACAGAGAGATCTTCAAGCTGTGCGCGAGGTGCGCGCGAGCTCCGAG 17251
QY 661 GTGCCGTTCTTCTCAACGCTGAGGAGCACTGCTGAGCACCACCACTCGACGCGCG 720
DB 17252 ATCCCGTCTTCTCAACGCTGACCGGCGCATGCTGCTGACACCACTGAGCGCGCG 17311
QY 721 TACTGTATCCGACACTGTGACAGCGCGTCCGTTTCAAGGATGCGGTCCAGGCGCTGCGG 780
DB 17312 TACTGTATCCGACACTGTGACAGCGCGTCCGTTTCAAGGATGCGGTCCAGGCGCTGCGG 17371
QY 781 GATGACGAGCAACCGGTCTTGTGAGAGTCAAGCCCCACCCACCTGCTCCCGCGCATC 840
DB 17372 GCGCGGACCACTGCTGCTTCAAGGATGAGCTCCACCCGCTGCTGCGCATGCTCGTC 17431
QY 841 GAGGACACCAACCGGAGCAACCGCGGAGAGCTCACCGCGATGCGGACGCTCCGCGCGCG 900
DB 17432 CAGGACATGATGAGAGCGCGGAGTGGGGGTGTGCGCTCGGACACCTGCGCGCGAG 17491
QY 901 GACGACGACACCGCGCGCTTC 921
DB 17492 AACGCGGTCTGACCGCTTC 17512

RESULT 12
AX211706
LOCUS AX211706 27541 bp DNA linear PAT 06-SBP-2001
DEFINITION Sequence 2 from Patent WO0159126.
ACCESSION AX211706
VERSION AX211706.1 GI:15523938
KEYWORDS
SOURCE
ORGANISM Streptomyces noursei.
Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 27541)
Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and
Gulliksen, O.M.
Gene cluster encoding a mycelatin polyketide synthase and its
manipulation and utility
Patent: WO 0159126-A 2 16-AUG-2001;
Norges Teknisk-Naturvitenskapelige Universitet (NO); STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOESKOLE (NO);
ALPHARMA AS (NO); SINVENT AS (NO); ZOTCHEV, SERGEY BORISOVICH
(NO); SEKUROVA, OLGA MIKOLAYEVNA (NO); FJAEVRIK, ESPEN (NO);
BRAUTASET, TRYGVE (NO); STROM, ARNE REIDAR (NO); VALLA, SVEIN

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz, L., Donadio, S. and McAlpine, J. B.
TITLE Method of directing biosynthesis of specific polyketides
JOURNAL Patent: US 6004787-A 3 21-DEC-1999;
FEATURES Location/Qualifiers
source 1..20235
/organism="unknown"
BASE COUNT 2534 a 7194 c 7767 g 2740 t
ORIGIN

Query Match 41.8%; Score 385.4; DB 6; Length 20235;
Best Local Similarity 65.6%; Pred. No. 1..1e-37;
Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;

Qy 1 GTCTTCGTGTTTCCCGACAGGCTGGGCAATGGGCCGGGATGGGTGTGCTGCTCGCC 60
Db 1696 GTCTTCCTGTTCCCGGGGAGGAGTTGCAATGGGCCGGGATGGGTGCGAGCTGTGTG 1755
Qy 61 TCCTCTCCGGTGTGTCGGCGGGAGTGCAGGCGGTGAGAGGCTCTGGCGCGGTGGTG 120
Db 1756 TCGTCGCGGTGTGTCGGCGGAGATCCGGGCTTGCAGCAGTCAGTGGCCCCGATGCAG 1815
Qy 121 GACTGGTCTGTGGTGGACATCTCTGCGCCGGGACGCGGGGATGCGGTGTGGAGCGGCGC 180
Db 1816 GACTGGAGAGTCTCCGACGTGTCTGCTCAGGCGCCGGGGGCGCGGAGCCGATGTC 1875
Qy 181 GATGTGTCAGCGTGTGCTGTTCAAGCTCATGTGTCTTTGTGCTGTGTGGCGTTCC 240
Db 1876 GACGTGTGTCAGCGGTGTGTTGTTCGCGGTATGTGTCTGTCGGGAGCTGTGTGCGCTCG 1935
Qy 241 TACGGTATGGAACCCGACCGGCTCTTGACCATTCACAGGCGGAGATCCGGCGCGGCGAT 300
Db 1936 TACGGGCTGAGACCCCGCGCGGTGTGCGGAGCACTTCAGAGGCGAGATCCCGCGCGCAC 1995
Qy 301 GTGTGTGGGGCGCTGAGCTGAAAGAGCGGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
Db 1996 GTTCGCGGGGCGCTCAAGTTGAGAGAGCGGGCGAAGCTGTGTGGGCGCGAGCGGCTG 2055
Qy 361 CTGGCGGCTGTGCGGGGCGGGGCGGCAATGGCTCTCAGTGCCTGCTGCTGCCAGAGATG 420
Db 2056 ATGCGGTGCTCTCCGGGAGGCGGCGCATGGCCGCTGCGGTGGGCGAGGCGGGGATG 2115
Qy 421 GAGCAGCTATTGTGTAGAGCGGTGGGCGGGCGGTTGTGTGTGCGGCGGTCAACGGCCCC 480
Db 2116 CCGGAGCGCTCGG--GCCGTGGCAGGACCGGCTCTCGGTGGCGCGGTCAACGGTCCC 2172
Qy 481 CGCTCCACCGCGCTCTCGGGGGAATGCCGAGCGGTGACGAGGTGCTGGGTACTGTGCGC 540
Db 2173 CGGTGCTGTGTGTTCTCCGCGAGCGGCGGCGGCGGCTGCGGCGTTTTCAGAGACTGCGCG 2232
Qy 541 GGCACCGGGGTGCGGGCCCGGCGGATCCGGTGCAGTATGCTCTGCGCACTGCCCCCATGTG 600
Db 2233 GCCGAGGGGATCGCGGTCCGCGCATGAGTGAATAGGACTAGCGGCTCGCGCAGATC 2292
Qy 601 CAGCCCTTGGGAGAGATTGCTGAGCTGTGGGGAATCAAGCCCGCAGCGGTCCGGC 660
Db 2293 GAGCGGGTCCGCGAGGAACTCTCCGAAAGACCGGCGACATCGCGCGCGCGCGCGCGCG 2352
Qy 661 GTGCGGTTCTTCTCAAGGTGAGGAGCACTGGCTGAGACCAACCTTGAACGCGGCC 720
Db 2353 GTGACGTTCACTCACTGTGAGTGGCGGTCTATGAGCGGACCGAGCTGATGCGCGG 2412
Qy 721 TACTGTATCCGAACTGTGACCAAGCGGTGCTGTTTCAAGCATGCTGTCCAGGCGCTGCGG 780
Db 2413 TACTGTATCCGAACTGTGCGGAGAGAGTGGCGGTTCGCCGAGCGCGGTGAGCGGGCTGCGG 2472
Qy 781 GATGAGGAGACCGCGGTCTTGTGAGATCAAGCCCGCACCACCTGCTGCTCCCGGATC 840
Db 2473 GAGTCGGGATACGACCGGTTCAATCGAGGTACAGCCCGCATTCGGGTCTGTGTCCAGGCGCGTC 2532

Qy 841 GAGAGACACCGAAGAGACCGCCGAGAGAGCTCACCGCGATCG 883
Db 2533 GAGAGGCGGTGGAAGAGGCTGACGCTGCCGAAGACGCGGTGCG 2575

Search completed: June 17, 2003, 17:55:32
Job time: 1791.01 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 183.212 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068

Sequence: 921 1 GCTCTGCTCTCCGAGCA.....ACAACGACCCGCTCTC 921

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N.GeneSeq_101002.*
1: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921	100.0	11916	22	AAH79279
2	921	100.0	12381	21	AA258381
3	921	100.0	30690	21	AA92301
4	921	100.0	30690	22	AAH79277
5	919.4	99.8	31432	21	AA92302
6	919.4	99.8	31432	22	AAH79278
7	402.4	43.7	65140	22	AAH7184
8	402.4	43.7	125401	22	AAH7186
9	400.8	43.5	27541	22	AAH7185

10	385.4	41.8	29879	14	AA046806
11	373.8	40.6	50937	21	AAA09469
12	362.8	39.4	1010	20	AAV99255
13	362.8	39.4	1010	21	AAA72521
14	351	38.1	53789	19	AAV21187
15	349.6	38.0	77788	22	AAH88335
16	349.6	38.0	50000	22	AAH88312
17	349.6	38.0	50000	22	AAH88316
18	348	37.8	80161	20	AAZ21501
19	347.4	37.7	47981	22	AAH30757
20	345	37.5	28958	17	AAH06769
21	345	37.5	28958	18	AAH89956
22	345	37.5	28958	21	AAH75299
23	345	37.5	49377	19	AAV05287
24	344.4	37.4	44377	18	AAH78508
25	344.4	37.4	44377	18	AAH80414
26	342.4	37.2	9513	22	AAH88337
27	342.4	37.2	50000	22	AAH88313
28	342.2	37.2	43280	18	AAH80413
29	340.4	37.0	16767	22	AAH88339
30	319	34.6	441529	22	AAH96682
31	318	34.5	13842	21	AAZ87297
32	318	34.5	36778	21	AAZ87318
33	318	34.5	37948	21	AAZ87285
34	318	34.5	38506	21	AAH75633
35	318	34.5	38506	21	AAZ56001
36	317.2	34.4	77536	21	AAH14651
37	311	33.8	11220	21	AAZ87298
38	308.6	33.5	1035	20	AAV99256
39	308.6	33.5	1035	21	AAH72538
40	308.6	33.5	4403765	22	AAH96683
41	306.8	33.3	4403765	22	AAH96683
42	306.8	33.3	441529	22	AAH96682
43	305.8	32.8	4689	21	AAZ87299
44	302	32.8	1881	24	AAH18436
45	298.4	32.4	68750	21	AAZ55887

ALIGNMENTS

RESULT 1					
ID	AAH79279	standard; DNA; 11916 BP.			
AC	AAH79279;				
XX					
DT	04-DEC-2001	(first entry)			
XX					
DE	Streptomyces avermectin coding sequence derivative	SEQ ID NO: 3.			
KW	Avermectin aglycone synthase; AAS; avermectin derivative;				
KW	drug production; veterinary drug; pesticide; da.				
XX					
OS	Synthetic.				
XX					
FT	Key	Location/Qualifiers			
FT	CDS	1..11916			
FT		/tag= a			
FT		/product= "AAG65268"			
XX		/partial			
XX					
XX	WO200162939-A1.				
XX					
XX	30-AUG-2001.				
XX					
XX	23-FEB-2001; 2001WO-0P01381.				
XX					
XX	24-FEB-2000; 2000JP-0047405.				
XX					
XX	(KYO) KYOWA HAKKO KOGYO KK.				
XX	(KITA) KITASATO INST.				
XX					

eryA region of S.
Streptococcus olea
DNA encoding an ac
DNA sequence of th
Amycolatopsis med
S. spinosa DNA fra
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
Microsporidia meg
Sorangium cellulos
DNA sequence of So
The soraphen biosy
Platenolide syntha
Platenolide syntha
S. spinosa DNA fra
S. spinosa DNA fra
Tylosin synthase
S. spinosa DNA fra
Mycobacterium tube
S. venezuelae macr
S. venezuelae macr
DNA encoding an ac
DNA encoding the A
Mycobacterium tube
Mycobacterium tube
Mycobacterium tube
S. venezuelae macr
Cortic 109 DNA enc
Sorangium cellulos

PI Endo H, Kamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
XX MPI: 2001-582053/65.
DR
XX P-PSDB; AAM65268.
PT
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
PS
XX
XX Claim 10; Page 149-167; 257bp; Japanese.
PS
XX
XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the *S. avermectilis* genome.
XX
XX
XX Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;
SQ

Query Match 100.0%; Score 921; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 1,66-146;
Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTCGTGTTTCCCGACAGGGTGGGAGTGGGCGGGAGTGGTGGTGGTCTCGCC 60
DB 3148 GTCTTCGTGTTTCCCGACAGGGTGGGAGTGGGCGGGAGTGGTGGTGGTCTCGCC 3207
QY 61 TCCTCTCCGGGTTCGCCCGGAGATGCAGGGGTGTCAGAGAGCTCTGGCCCGTGGGTG 120
DB 3208 TCCTCTCCGGGTTCGCCCGGAGATGCAGGGGTGTCAGAGAGCTCTGGCCCGTGGGTG 3267
QY 121 GACTGATCTGTGTGAGACATCTCTGGCGCGGAGACGCGGGGAGATGCGTGTGGAGAGCGGGCC 180
DB 3268 GACTGATCTGTGTGAGACATCTCTGGCGCGGAGACGCGGGGAGATGCGTGTGGAGAGCGGGCC 3327
QY 181 GATGTGTTCACAGCTGTGTGCTTTCAAGCGTCATGATGTCTTTGGCTCTGTGGCGTTCC 240
DB 3328 GATGTGTTCACAGCTGTGTGCTTTCAAGCGTCATGATGTCTTTGGCTCTGTGGCGTTCC 3387
QY 241 TACGGTATCGAACCCGAGCGGCTCTTGCCCATTTCCAGAGGCGAGATGCGGCGCGCAT 300
DB 3388 TACGGTATCGAACCCGAGCGGCTCTTGCCCATTTCCAGAGGCGAGATGCGGCGCGCAT 3447
QY 301 GTGTGTGGGGGCGCTAGGCTGAAAGACGGGCGAGACATGTTGGCTGGCGACCGCGGGCG 360
DB 3448 GTGTGTGGGGGCGCTAGGCTGAAAGACGGGCGAGACATGTTGGCTGGCGACCGCGGGCG 3507
QY 361 CTGGCGCGTGTGCGGGGCGCGGGCGGAGATGAGCTAGTGCCTGCTCTGACCAGAGAGTG 420
DB 3508 CTGGCGCGTGTGCGGGGCGCGGGCGGAGATGAGCTAGTGCCTGCTCTGACCAGAGAGTG 3567
QY 421 GAGCAGCTCATTTGTGAAGCGGTGGCGCGGGCGGTTGTGGGTGGCGGCGGTCAACGCGCC 480
DB 3568 GAGCAGCTCATTTGTGAAGCGGTGGCGCGGGCGGTTGTGGGTGGCGGCGGTCAACGCGCC 3627
QY 481 CGCTCCACCGCGCTTCGGGGGGGATGCCAGACGGGTGACAGAGTGTGGCGGTCACTGTGCC 540
DB 3628 CGCTCCACCGCGCTTCGGGGGGGATGCCAGACGGGTGACAGAGTGTGGCGGTCACTGTGCC 3687
QY 541 GGCAACGGGGGTGCGGGGCGCGGCGAGATCCCGATGCACTATGCTTGSCATGCCCCCATGTG 600
DB 3688 GGCAACGGGGGTGCGGGGCGCGGCGAGATCCCGATGCACTATGCTTGSCATGCCCCCATGTG 3747
QY 601 CAGCCCTTGCGGAGAGAGTGTCTGAGCTGTGGGGAATCAGCCCGCAGCCGTCGCGC 660
DB 3748 CAGCCCTTGCGGAGAGAGTGTGTGAGCTGTGGGGAATCAGCCCGCAGCCGTCGCGC 3807
QY 661 GTGCCGTTCTTCTCAAGGTGAGGGGCACTGGCTGACACCAACCTTGAGCGCGGCC 720

Query Match	Score	DB	Length
100.0%	921	DB 21	13381


```

QY 241 TACGATATGAAACCCGACGCGGTCTTGCCATTCCAGGAGCGAGATCGCGCGGCAT 300
DB 3388 TAGGATATGAAACCCGACGCGGTCTTGCCATTCCAGGAGCGAGATCGCGCGGCAT 3447
QY 301 GTGTGTGGGGCGCTGAGCTTGAAGACGCGCGCGAAGACTGTTCGCTGCGACCGCGCG 360
DB 3448 GTGTGTGGGGCGCTGAGCTTGAAGACGCGCGCGAAGACTGTTCGCTGCGACCGCGCG 3507
QY 361 CTGGCGCGCTGTGGGGCGCGCGCGCGAGTCCATGATCGCGCTGCTGCGCGAGAGTG 420
DB 3508 CTGGCGCGCTGTGGGGCGCGCGCGCGAGTCCATGATCGCGCTGCTGCGCGAGAGTG 3567
QY 421 GAGCAGCTCATTTGTGAGCGGTGGGCGCGCGCTGTGGGTGCGCGCGCTCAACGAGCC 480
DB 3568 GAGCAGCTCATTTGTGAGCGGTGGGCGCGCGCTGTGGGTGCGCGCGCTCAACGAGCC 3627
QY 481 CGCTCCACCGCGCTCTCGGGGAGTGCAGAGCGGTGAGACGAGTGTGCGTACTGTGCC 540
DB 3628 CGCTCCACCGCGCTCTCGGGGAGTGCAGAGCGGTGAGACGAGTGTGCGTACTGTGCC 3687
QY 541 GGCACCGGGGTGGGGCGCGCGCGAGTCCCGGTGCACTATGCTTCCGACTGCCCATGTG 600
DB 3688 GGCACCGGGGTGGGGCGCGCGCGAGTCCCGGTGCACTATGCTTCCGACTGCCCATGTG 3747
QY 601 CAGCCCTCGGGAGAGATTGTGAGCTGTGGGGAGCATCAGCCCGAGCGCGCTCGGC 660
DB 3748 CAGCCCTCGGGAGAGATTGTGAGCTGTGGGGAGCATCAGCCCGAGCGCGCTCGGC 3807
QY 661 GTGCCGTTCTTCTCAACGCTGAGGGGACCTGTGAGACACCAACCTTGACGCGCC 720
DB 3808 GTGCCGTTCTTCTCAACGCTGAGGGGACCTGTGAGACACCAACCTTGACGCGCC 3867
QY 721 TACTGTATCCGGAACCTGACACGCGGTCCGTTTACAGGATCCGCTCAGGCGCTGGG 780
DB 3868 TACTGTATCCGGAACCTGACACGCGGTCCGTTTACAGGATCCGCTCAGGCGCTGGG 3927
QY 781 GATGACGACACCGCGCTTCTGTGAAGTCAAGCCCGACCGCTGTCGCCGCGCATC 840
DB 3928 GATGACGACACCGCGCTTCTGTGAAGTCAAGCCCGACCGCTGTCGCCGCGCATC 3987
QY 841 GAAAGACACACCGAAGACACCGCGAAGAGTCAACCGCGATCGGCGCGCGCG 900
DB 3988 GAAAGACACACCGAAGACACCGCGAAGAGTCAACCGCGATCGGCGCGCGCG 4047
QY 901 GACAAAGACACCGCGCGCTTC 921
DB 4048 GACAAAGACACCGCGCGCTTC 4068

RESULT 4
AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
AC AAH79277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermectilis coding sequences SEQ ID NO: 1.
XX
KM Avermectin aglycone synthase; AAS; avermectin derivative;
XX
KW drug production; veterinary drug; pesticide; ds.
XX
OS Streptomyces avermectilis.
XX
XX
XX Key Location/Qualifiers
XX FT 1..11919
XX FT /tag= a
XX FT /product= "AAG65264"
XX FT /partial
XX FT 11971..30690
XX CDS
XX FT /tag= b
XX FT /product= "AAG65265"

```

```

XX
PN MO200162939-A1.
XX
PD 30-AUG-2001.
XX
XX
XX PF 23-FEB-2001; 2001WO-JP01381.
XX
XX PR 24-FEB-2000; 2000JP-0047405.
XX
XX PA (KYOMA ) KYOMA HAKKO KOGYO KK.
XX (KITA ) KITASATO INST.
XX
XX PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
XX DR MPI: 2001-582053/65.
XX P-PsDB; AAG65264, AAG65265.
XX
XX PT New modified avermectin aglycone synthase derived from Streptomyces
XX PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
XX PT drugs and pesticides -
XX
XX PS Example 2; Page 58-123; 257pp; Japanese.
XX
XX CC The present invention relates to the production of modified derivatives
XX CC of avermectin aglycone synthase (AAS) derived from Streptomyces
XX CC avermectilis. The activity of an acyl carrier protein (ACP),
XX CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
XX CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX CC suppressed. The process can be used in the production of drugs, veterinary
XX CC drugs and pesticides. The present sequence is a fragment of the S.
XX CC avermectilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 921; DB 22; Length 30690;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-146;
XX Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GTCTTCGCTTCCCGGACAGGAGTGGGAGAGTGGCGGGAGATGGGTGCTGCTCC 60
DB 3148 GTCTTCGCTTCCCGGACAGGAGTGGGAGAGTGGCGGGAGATGGGTGCTGCTCC 3207
QY 61 TCCTCTCGGATGTTGCGCGCGGAGATGACAGGAGCTCTGAGCGCGTGGTG 120
DB 3208 TCCTCTCGGATGTTGCGCGCGGAGATGACAGGAGCTCTGAGCGCGTGGTG 3267
QY 121 GACTGCTGTGTGGAATCTCTGCGCGGAGCGCGGGAGTGGGTGGAGCGGCGC 180
DB 3268 GACTGCTGTGTGGAATCTCTGCGCGGAGCGCGGGAGTGGGTGGAGCGGCGC 3327
QY 181 GATGTGTCCAGACCTGTCTGTTCAGCGTCAATGATGTTTGGCGTCTGTGGGTTCC 240
DB 3328 GATGTGTCCAGACCTGTCTGTTCAGCGTCAATGATGTTTGGCGTCTGTGGGTTCC 3387
QY 241 TACGATATGAAACCCGACGCGGTCTTGCCATTCCAGGAGCGAGATCGCGCGCAT 300
DB 3388 TACGATATGAAACCCGACGCGGTCTTGCCATTCCAGGAGCGAGATCGCGCGCAT 3447
QY 301 GTGTGTGGGGCGCTGAGCTTGAAGACGCGCGCGAAGACTGTTCGCTGCGACCGCGCG 360
DB 3448 GTGTGTGGGGCGCTGAGCTTGAAGACGCGCGCGAAGACTGTTCGCTGCGACCGCGCG 3507
QY 361 CTGGCGCGCTGTGGGGCGCGCGCGCGAGTCCATGATCGCGCTGCTGCGCGAGAGTG 420
DB 3508 CTGGCGCGCTGTGGGGCGCGCGCGCGAGTCCATGATCGCGCTGCTGCGCGAGAGTG 3567
QY 421 GAGCAGCTCATTTGTGAGCGGTGGGCGCGCGCTGTGGGTGCGCGCGCTCAACGAGCC 480
DB 3568 GAGCAGCTCATTTGTGAGCGGTGGGCGCGCGCTGTGGGTGCGCGCGCTCAACGAGCC 3627
QY 481 CGCTCCACCGCGCTCTCGGGGAGTGCAGAGCGGTGAGACGAGTGTGCGTACTGTGCC 540

```

Db 3628 CGCTCACCGCCCTCTCGGGGGATGCCAGGCGGTGACGAGTGTGGCTATCTGTGCC 3687
Qy 541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTGACTATGCTCGACTGCCCCCATGTG 600
Db 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTGACTATGCTCGACTGCCCCCATGTG 3747
Qy 601 CAGCCCTTGGGAGAGATTGCTGAGCTGCTGGGGACATCAGCCCGCAGCCGTCCGCG 660
Db 3748 CAGCCCTTGGGAGAGATTGCTGAGCTGCTGGGGACATCAGCCCGCAGCCGTCCGCG 3807
Qy 661 GTCGCGTCTTCTCCACCGGTGAGGGGACCTGTGCTGAGACCAACCACTCGAGCGCGGC 720
Db 3808 GTGCCCTTCTTCTCCACCGGTGAGGGGACCTGTGCTGAGACCAACCACTCGAGCGCGGC 3867
Qy 721 TACTGTACCGCAACCTGACACGCGGTCCGTTTACGAGTACCGTCCAGGCGCTTGCGG 780
Db 3868 TACTGTACCGCAACCTGACACGCGGTCCGTTTACGAGTACCGTCCAGGCGCTTGCGG 3927
Qy 781 GATGACGACACCGCGTCTTCTGTCGAACTCAGCCCGCACCCTCTGTCGCCCATC 840
Db 3928 GATGACGACACCGCGTCTTCTGTCGAACTCAGCCCGCACCCTCTGTCGCCCATC 3987
Qy 841 GAAGACACCAACCGGAAACCGCGGAAAGCTGACCGCGATCGGCGAGCTTCGCGCGCGC 900
Db 3988 GAAGACACCAACCGGAAACCGCGGAAAGCTGACCGCGATCGGCGAGCTTCGCGCGCGC 4047
Qy 901 GACCAACGACACCGCGCGCTTC 921
Db 4048 GACCAACGACACCGCGCGCTTC 4068

RESULT 5
AAA92302
ID AAA92302 standard; DNA; 31422 BP.

AC AAA92302;
DT 10-JAN-2001 (first entry)
XX S. avermectin aglycon synthase DNA avail SEQ ID NO:2.
XX Streptomyces avermectin aglycon synthase; biosynthesis;
XX multi-functional enzyme; polyketide; avermectin; veterinary drug;
XX agrochemical; ds.
OS Streptomyces avermectin.
XX
XX Key Location/Qualifiers
FH 1..14646
FT /tag= a
FT /note= "Avermectin aglycon synthase protein"
FT 14824..31422
FT /tag= b
FT /note= "Avermectin aglycon synthase protein"
XX
XX MO200050605-A1.
XX
XX 31-AUG-2000.
XX
XX 23-FEB-2000; 2000MO-JP01041.
XX
XX 24-FEB-1999; 99JP-0046961.
XX
XX (KITA) KITASATO INST.
XX
XX Omura S, Ikeda H;
XX
XX MPI; 2000-565458/52.
XX
XX P-PDB; AAB3751, AAB3752.
XX
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use

XX
PS Claim 2; Page 134-203; 314pp; Japanese.
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of avermectin aglycon
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5311 T; 0 other;

Query Match 99.8%; Score 919.4; DB 21; Length 31422;
Best Local Similarity 99.8%; Pred. No. 2,9e-146;
Matches 920; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCTCGATGTTCCCGGACAGGGTGGGAGTGGGCGGGATGGGTGCTGCTCGCC 60
Db 16630 GTCTCGATGTTCCCGGACAGGGTGGGAGTGGGCGGGATGGGTGCTGCTCGCC 16689
Qy 61 TCTCTCCGATGTTGCGCCGCGAGTGCAGCGGTGCGAGAGGCTTGCGCGCGTGGTG 120
Db 16690 TCTCTCCGATGTTGCGCCGCGAGTGCAGCGGTGCGAGAGGCTTGCGCGCGTGGTG 16749
Qy 121 GACTGATGTTGAGTGAATCTGTGCGCGGAGCGCGGGAGTGCAGTGTGGAGCGGAGC 180
Db 16750 GACTGATGTTGAGTGAATCTGTGCGCGGAGCGCGGGAGTGCAGTGTGGAGCGGAGC 16809
Qy 181 GATGTGTCCACCGCTGTGCTTTCAGCGTCAAGTGTCTTGTGGCTGTGCGCTTCC 240
Db 16810 GATGTGTCCACCGCTGTGCTTTCAGCGTCAAGTGTCTTGTGGCTGTGCGCTTCC 16869
Qy 241 TACGATATGAACCGGACGCGGTCTTGGCACTTCCAGGCGAGATCGCGCGCGCAT 300
Db 16870 TACGATATGAACCGGACGCGGTCTTGGCACTTCCAGGCGAGATCGCGCGCGCAT 16929
Qy 301 GTGTGTGGGCGCTGAGCTGAAAGACCGGCGGAAAGCTTGTGCTGTGCGACCGCGCG 360
Db 16930 GTGTGTGGGCGCTGAGCTGAAAGACCGGCGGAAAGCTTGTGCTGTGCGACCGCGCG 16989
Qy 361 CTGGCGCGCTGTGCGGGCGCGGCGGATGCGCTCAAGTCCCTGCTGCCAGAGATG 420
Db 16990 CTGGCGCGCTGTGCGGGCGCGGCGGATGCGCTCAAGTCCCTGCTGCCAGAGATG 17049
Qy 421 GAGCAGCTCATTTGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAAACGGCCCC 480
Db 17050 GAGCAGCTCATTTGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAAACGGCCCC 17109
Qy 481 CGCTCACCGCGGTCTCGGGGATGCCAGCGGTGAGCAGAGTGTCTGTACTGTGCG 540
Db 17110 CGCTCACCGCGGTCTCGGGGATGCCAGCGGTGAGCAGAGTGTCTGTACTGTGCG 17169
Qy 541 GGCACCGGGGTGCGGGCGCGGATGCCAGTGTGACTATGCTCGCATGCGCCCATGTG 600
Db 17170 GGCACCGGGGTGCGGGCGCGGATGCCAGTGTGACTATGCTCGCATGCGCCCATGTG 17229
Qy 601 CAGCCCTTGGGAGAGATTGCTGAGCTGCTGGGAGGACATCAGCCCGCAGCGGTCCGCG 660
Db 17230 CAGCCCTTGGGAGAGATTGCTGAGCTGCTGGGAGGACATCAGCCCGCAGCGGTCCGCG 17289
Qy 661 GTGCCCTTCTTCTCCACCGGTGAGGCACTGTGTGACCAACCACTCTGACGCGCGC 720
Db 17290 GTGCCCTTCTTCTCCACCGGTGAGGCACTGTGTGACCAACCACTCTGACGCGCGC 17349
Qy 721 TACTGTACCGCAACCTGACACGCGGTCCGTTTACGAGTACCGTCCAGGCGCTTGCGG 780
Db 17350 TACTGTACCGCAACCTGACACGCGGTCCGTTTACGAGTACCGTCCAGGCGCTTGCGG 17409

QY 781 GATGACGACACCGCGTCTTCTGTCGAAAGTCAGCCCCACCCCTCGTCCCGGCATC 840
DB 17410 GATGACGACACCGCGTCTTCTGTCGAAAGTCAGCCCCACCCCTCGTCCCGGCATC 17469
QY 841 GAAAGACACACCGGAGACACCGCCGAAAGACGTACCGCGATCGGACGCTCGCGCGGC 900
DB 17470 GAAAGACACACCGGAGACACCGCCGAAAGACGTACCGCGATCGGACGCTCGCGCGGC 17529
QY 901 GACACGACACCGCGCGCTTC 921
DB 17530 GACACGACACCGCGCGCTTC 17550

RESULT 6
AAH79278
ID AAH79278 standard; DNA; 31422 BP.

AC AAH79278;
XX
DT 04-DEC-2001 (first entry)
XX

DE Streptomyces avermectilis coding sequences SEQ ID NO: 2.
XX
KM Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide; ds.
OS Streptomyces avermectilis.

XX
FH Key Location/Qualifiers
FT 1..1466
FT /*tag= a
FT /*product= "AAG65266"
FT 14824..31422
FT /*tag= b
FT /*product= "AAG65267"

XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
XX P-PSDB; AAG65266, AAG65267.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
XX avermectilis used in production of 22,23-dihydrovermectin B1a used in
XX drugs and pesticides -
XX
PS Disclosure; Page 103-149; 257pp; Japanese.

XX
CC The present invention relates to the production of modified derivatives
XX of avermectin aglycone synthase (AAS) derived from Streptomyces
XX avermectilis. The activity of an acyl carrier protein (ACP),
XX beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
XX beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX suppressed. The process can be used in the production of drugs, veterinary
XX drugs and pesticides. The present sequence is a fragment of the S.
XX avermectilis genome.

XX
SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
XX
XX Query Match 99.8%; Score 919.4; DB 22; Length 31422;
XX Best Local Similarity 99.9%; Pred. No. 2.9e-146;
XX Matches 920; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCTTCGTGTTTCCGGAACAGGGGTGGAGTGGGCGGGATGGGTGTGCTGTGCTGCC 60
DB 16630 GTCTTCGTGTTTCCGGAACAGGGGTGGAGTGGGCGGGATGGGTGTGCTGTGCTGCC 16689
QY 61 TCCTTCGCGGTGTTCCCGCGGAGATGACAGCGTGCAGAGAGGCTCTGCGCGCTGGGTG 120
DB 16690 TCCTTCGCGGTGTTCCCGCGGAGATGACAGCGTGCAGAGAGGCTCTGCGCGCTGGGTG 16749
QY 121 GACTGATCTGTGTGGACATCTCTGCGCGGGAGATGCGGTGTGGAGAGCGGGCC 180
DB 16750 GACTGATCTGTGTGGACATCTCTGCGCGGGAGATGCGGTGTGGAGAGCGGGCC 16809
QY 181 GATGTGTCCAGCCTGTGTGCTGTTCAAGCGTCATGTGTCTTGGTGTCTGTGGCGTTCC 240
DB 16810 GATGTGTCCAGCCTGTGTGCTGTTCAAGCGTCATGTGTCTTGGTGTCTGTGGCGTTCC 16869
QY 241 TACGGTATGAAACCCGACCGCGGTCTTGGCCATTCCAGGCGGAGATGCGCGCGCAT 300
DB 16870 TACGGTATGAAACCCGACCGCGGTCTTGGCCATTCCAGGCGGAGATGCGCGCGCAT 16929
QY 301 GTGTGTGGGGCGCTGAGCTGAAAGACCGCGGAGAACTGTGTGCGCTGGCGCGGGCG 360
DB 16930 GTGTGTGGGGCGCTGAGCTGAAAGACCGCGGAGAACTGTGTGCGCTGGCGCGGGCG 16989
QY 361 CTGGCGCGTGTGCGGGGCGCGGCGGCAATGAGCTGAGTGCCTGCTGCCAGGAGGTG 420
DB 16990 CTGGCGCGTGTGCGGGGCGCGGCGGCAATGAGCTGAGTGCCTGCTGCCAGGAGGTG 17049
QY 421 GAGCAGCTGATGTGTGAGCGGTGTGGCGCGGTGTGTGGGTGGCGCGGTCAACGCGCCC 480
DB 17050 GAGCAGCTGATGTGTGAGCGGTGTGGCGCGGTGTGTGGGTGGCGCGGTCAACGCGCCC 17109
QY 481 CGCTCCACCGCGCTCTCGGGGGATGCCAGGCGGTGTGACGAGTCTGGCGCTGATGTGTC 540
DB 17110 CGCTCCACCGCGCTCTCGGGGGATGCCAGGCGGTGTGACGAGTCTGGCGCTGATGTGTC 17169
QY 541 GGCACCGGGGTGCGGGGCGCGGCGGATCCCGGTGCACTATGCTCGCACTGCCCCATGTG 600
DB 17170 GGCACCGGGGTGCGGGGCGCGGCGGATCCCGGTGCACTATGCTCGCACTGCCCCATGTG 17229
QY 601 CAGCCCTGCGGGAGAGTGTGTGAGCTGTGTGGGGACATCAGCCGCGAGCGTCCGGC 660
DB 17230 CAGCCCTGCGGGAGAGTGTGTGAGCTGTGTGGGGACATCAGCCGCGAGCGTCCGGC 17289
QY 661 GTGCCGTTTCTTCCACGCTGTGAGGCGACCTGAGTGTGACACCAACCTTGAACCGCGCC 720
DB 17290 GTGCCGTTTCTTCCACGCTGTGAGGCGACCTGAGTGTGACACCAACCTTGAACCGCGCC 17349
QY 721 TACTGGTACCGCAACCTGACACGCGGTGCGTTTCAAGATGCGGTCAGGCGCTTGAGC 780
DB 17350 TACTGGTACCGCAACCTGACACGCGGTGCGTTTCAAGATGCGGTCAGGCGCTTGAGC 17409
QY 781 GATGACGACACCGCGGTCTTGTGTGAGTCAAGCTCAAGCCCAACCCCTGTGTCCCGCATC 840
DB 17410 GATGACGACACCGCGGTCTTGTGTGAGTCAAGCTCAAGCCCAACCCCTGTGTCCCGCATC 17469
QY 841 GAAAGACACACCGGAGACACCGCCGAAAGACGTACCGCGATCGGACGCTCGCGCGGC 900
DB 17470 GAAAGACACACCGGAGACACCGCCGAAAGACGTACCGCGATCGGACGCTCGCGCGGC 17529
QY 901 GACACGACACCGCGCGCTTC 921
DB 17530 GACACGACACCGCGCGCTTC 17550

RESULT 7
AAD17184
ID AAD17184 standard; DNA; 65140 BP.

AC AAD17184;
XX
DT 29-NOV-2001 (first entry)

XX Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
 DE Polyketide synthase; PKS; macroide; nystatin; PKS gene cluster;
 XX antifungal; antibiotic; nysI; de.
 XX Streptomyces noursei.
 OS
 XX Key Location/Qualifiers
 FH complement (1..1035)
 PT /tag= a
 PT /product= "NysD2 partial protein"
 PT /note= "CDS does not include stop codon"
 PT complement (1056..2576)
 PT /tag= b
 PT /product= "NysD1 protein"
 PT 2806..6906
 PT /tag= c
 PT /product= "NysA protein"
 PT 6952..16530
 PT /tag= d
 PT /product= "NysB protein"
 PT 16550..49840
 PT /tag= e
 PT /product= "NysC protein"
 PT 50260..51015
 PT /tag= f
 PT /product= "NysE protein"
 PT 51405..54305
 PT /tag= g
 PT /product= "NysR1 protein"
 PT 54329..57190
 PT /tag= h
 PT /product= "NysR2 protein"
 PT /note= "CDS does not include start codon"
 PT 57180..59963
 PT /tag= i
 PT /product= "NysR3 protein"
 PT 60415..61047
 PT /tag= j
 PT /product= "NysR4 (short) protein"
 PT /note= "CDS does not include start codon"
 PT 61736..62497
 PT /tag= k
 PT /product= "NysR5 protein"
 PT /note= "CDS does not include start codon"
 PT complement (62551..63615)
 PT /tag= l
 PT /product= "ORF2 protein"
 PT /note= "CDS does not include start codon"
 PT 63765..64961
 PT /tag= m
 PT /product= "ORF1 protein"
 PT
 XX
 XX WO200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 XX 08-FEB-2001; 2001WO-GB0509.
 XX
 XX 08-FEB-2000; 2000GB-0002840.
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX
 PA (UNO-) UNIV NORGE TEKNIKTUTSKAPBELLIGE.
 PA (SNT) SINTER STIFTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SIN-) SINVENT AS.
 PA (DZIE-) DZIELEWSKA H.
 PA (ZOTC-) ZOTCHEV S B.
 PA (SEKU-) SEKUROVA O N.
 PA (FJAE-) FJAEVIK E.
 PA (BRAU-) BRAUTASET T.

PA (STRO/) STROM A R.
 XX
 XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 XX WPI; 2001-557614/62.
 DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
 DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides.
 PT useful as antibiotics and antifungals -
 XX
 XX Claim 2; Page 116-151; 266bp; English.
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macroide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
 XX
 SQ Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
 Query Match 43.7%; Score 402.4; DB 22; Length 65140;
 Best Local Similarity 66.2%; Pred. No. 2,3e-59;
 Matches 612; Conservative 0; Mismatches 306; Indels 6; Gaps 2;
 QY 1 GTCTCGTGTTCCTCCGACAGAGGTGGGCGAGTGGGATGGTGTGCTGTCTCCGCC 60
 DB 8674 GTGTTCGTATTCCTCCGACAGAGGTGGGCGAGTGGGATGGTGTGCTGTCTCCGCC 8733
 QY 61 TCTCTCCGATGTTCCTCCGACAGAGGTGGGCGAGTGGGATGGTGTGCTGTCTCCGCC 120
 DB 8734 GAGTCGCGGTGTTCCTCCGACAGAGGTGGGCGAGTGGGATGGTGTGCTGTCTCCGCC 8793
 QY 121 GACTGCTGTGTGTGACATCTTCGCGCGGACGCGGAGATGGTGTGAGCGCGCC 180
 DB 8794 GACTGCTGTGTGTGACATCTTCGCGCGGAGATGGTGTGAGCGCGCC 8853
 QY 181 GATGT 240
 DB 8854 GATGT 8913
 QY 241 TAGGTATGAAACCCGACGCGGTCTTCGACATTCGACGCGGAGATGGTGTGCTGTCTCCGCC 300
 DB 8914 CGT 8973
 QY 301 GT 360
 DB 8974 GT 9033
 QY 361 CTGG--CCGT 417
 DB 9034 ATTGT 9093
 QY 418 GTGAGCAGCTCATTTGT 477
 DB 9094 CTGGAACCGCGT 9150
 QY 478 CCCCGCTCACCGCGGT 537
 DB 9151 CCGCGCTCGT 9210
 QY 538 GCCGCGACCGGAGGT 597
 DB 9211 ACCGCGACGATCCGCGCGCGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9270
 QY 598 GTGACGCCCTCGGAGAGGT 657
 DB 9271 GTGAGAGCTGT 9330
 QY 658 GCGGTGCGGT 717
 DB 9331 GAGGTGCGGT 9390

QY	718	GCCTACTGTATACCGGAACCTGGACACGAGCCGGTTCGAGGATGACCGTCCAGGCCCTG	7718
Db	9331	GGCTACTGTATTCGGAACTCTGGCGGACGGGTGCGGTTTCGCGAACCGATGCGGACCTTG	9450
QY	778	GCGATGACGGAACACCGCGTCTTCTGTGAATGACGCCCCACCCACCTGATCCCGCC	837
Db	9451	CTGGCGGCGGAGTATCCGCGCGCTTCGTCAGAGTCACTGCACCCGGTCTGCATGAGTCG	9510
QY	838	ATCGAAGACACACCGGAGAACACCGCCGGAAGACGTACCGGATGCGACGCTCCGCGCC	897
Db	9511	GTCTTGACCTGATGAGAGAGCCCGGGTCAAGCCGCTGCGACCGGACCTCGCGCGT	9570
QY	898	GGCGACAACGACACCCCGCGCTTC	921
Db	9571	GACCAAGGTGCGCGCGCGCTTC	9594
RESULT 8			
AD	AAD17186	standard; DNA; 125401 BP.	
XX	AAD17186;		
XX	29-NOV-2001	(first entry)	
XX	Streptomyces noursei	PKS gene cluster DNA.	
XX	Polysaccharide synthase; PKS; macrolide; nystatin; PKS gene cluster;		
XX	antifungal; antibiotic; ds.		
XX	Streptomyces noursei.		
XX	Key	Location/Qualifiers	
XX	FT CDS	6337..34771	
XX	FT	/*tag= a	
XX	FT	/product= "NysI complete protein"	
XX	FT	34792..51099	
XX	FT	/*tag= b	
XX	FT	/product= "NysJ protein"	
XX	FT	51155..57355	
XX	FT	/*tag= c	
XX	FT	/product= "NysK protein"	
XX	FT	57503..58687	
XX	FT	/*tag= d	
XX	FT	/product= "NysL protein"	
XX	FT	complement (58786..58980)	
XX	FT	/*tag= e	
XX	FT	/product= "NysM protein"	
XX	FT	/note= "CDS does not include start codon"	
XX	FT	complement (59045..60241)	
XX	FT	/*tag= f	
XX	FT	/product= "NysN protein"	
XX	FT	/note= "CDS does not include start codon"	
XX	FT	complement (60238..61296)	
XX	FT	/*tag= g	
XX	FT	/product= "NysO2 complete protein"	
XX	FT	120628..121308	
XX	FT	/*tag= h	
XX	FT	/product= "NysR4 (long) protein"	
XX	PA	(UNO-) UNIV NORGES TEKNISK NATURVITENSKAPLIGE.	
XX	PA	(SNTF) SINTEF STIFTELSEN IND TEK FORSK.	

PA	(ALPH-) ALPHARMA AS.
PA	(STIN-) SINVENT AS.
PA	(DZIE/) DZIBELSKA H.
PA	(ZOTC/) ZOTCHEV S B.
PA	(SEKU/) SEKUROVA O N.
PA	(FJAE/) FJAERVIK E.
PA	(BRAU/) BRAUTASET T.
PA	(STRQ/) STROM A R.
XX	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;
DR	WPI: 2001-557614/62.
DR	P-PDSB; AAEI0143, AAEI0144, AAEI0145, AAEI0146, AAEI0147, AAEI0148,
DR	AAEI0149, AAEI0150.
XX	New nystatin polyketide synthase polynucleotides and polypeptides,
PT	useful as antibiotics and antifungals -
XX	Claim 1; Page 188-254; 266pp; English.
PS	
CC	The present invention relates to the cloning and sequencing of the gene
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC	involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC	The nystatin PKS is useful as antifungal antibiotics. The present
CC	sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX	
SQ	Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
	Query Match 43.7%; Score 402.4; DB 22; Length 125401;
	Best Local Similarity 66.2%; Pred. No. 2.2e-59;
	Matches 612; Conservative 0; Mismatches 306; Indels 6; Gaps 2;
QY	1 GTCTTCGATGTTTCCCGACAGGGTGCGAGTAGTGGCCGGAGATGTCGTGCTGCC 60
DB	68935 GTGTTCGTGTTCCCGGTACAGGTTTCGAGTGGTGGGAATGGGGCCCACTCTGCAT 68994
QY	61 TCCTCTCCGGTGTTCGCCCGCGGATGCAAGCGGTGCAGAGAGCTTGGCGCCGTGGTG 120
DB	68995 GACTCGCGCGGTGTTCCGAGAGCGGATGCGAGTGTGCGCGGCACTCCCGATTCAACC 69054
QY	121 GACTGGTGTGTGTGGACATCCCTGCGCCCGGACGCGGGGATGCGTGTGGAGAGCGGCCC 180
DB	69055 GACTGTGTCGTGTGATGATGCTGCTCGGGGTGTGTGTGGTGGCGCCCGTGTGGAGCGGGTC 69114
QY	181 GATGTGGTCCAGCCTTGCTGTGTTCAGCTCATGATGATGTTGTGGCTGTGTGACGTTCC 240
DB	69115 GATGTGTGTGACGCGGCGCTGTTCCGGGTGATGTGTCTGTGTGGCTGTGTGGGGTTCC 69174
QY	241 TACGTTATCGAACCCGACGCGGTCTTTGGCAATTCGACAGAGAGATCGCGGCGCGCAT 300
DB	69175 CGTGTGTGTGTGCGGATGCGGATGCGGTGTGGAGCATTCGACAGGTGATGCTGCCGCGGTG 69234
QY	301 GTGTGTGGGGCGCTGAGCCTGAAGAAGCGCGCGAAGACTGTGGCTGTGGACCGGGGCG 360
DB	69235 GTGTTCGAGTGCCTGTGCTTTGGAGGACGCGGGCGCGGGTGTGGCCTGTGGAGTCAAGCC 69294
QY	361 CTGG---CCGCTGTGTGGGGCGCGGGCGGACATGAGCCTCAAGTGCCTGCTCCAGAG 417
DB	69295 ATTGTGTGTGCTGTGGCGGGGCGGGCGGGAGTAATGTCCGTGCGCTGTGTGGTGAAGCTG 69354
QY	418 GTTGAGCACTATTGTGTAGCGGTGTGGCGGGCGGGTGTGTGGTGTGAGCGAGCTCAACGAC 477
DB	69355 CTGGAACCGCGGTGTGCA---GTTTCAAGGGGCGGGGTGTGTGTGTGGTGTGCAACGGC 69411
QY	478 CCCCGCTTCAACGCGCGTCTTCGGGGGATGCGAGCGGTGTGACAGAGTGTGGCGTAACTGT 537
DB	69412 CCGCGCTCTGTGTGTGTTCGCCGCAAGCCCGAGGCGGTGTGACGCGCTGACGCCCCGCTG 69471
QY	538 GCCGGCACCGGGGTGTGGGCGCGGGCGGAGATCCGGGTGTGACTATGCTGTGACATGCGCCCAT 597
DB	69472 ACCGCGACGACATTCGGGCGCGCGGATCGCGGTGTGACTACGCTTCGACTTCGACCAAG 69531

Accession	Gene	Species	Protein	Length	Score	E-value	Bit Score	Identities	Positives	Gaps	Conserved Domains	Annotations
U00001	trpA	Escherichia coli	tryptophan synthase alpha subunit	401	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase alpha subunit
U00001	trpB	Escherichia coli	tryptophan synthase beta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase beta subunit
U00001	trpC	Escherichia coli	tryptophan synthase gamma subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase gamma subunit
U00001	trpD	Escherichia coli	tryptophan synthase delta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase delta subunit
U00001	trpE	Escherichia coli	tryptophan synthase epsilon subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase epsilon subunit
U00001	trpF	Escherichia coli	tryptophan synthase zeta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase zeta subunit
U00001	trpG	Escherichia coli	tryptophan synthase eta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase eta subunit
U00001	trpH	Escherichia coli	tryptophan synthase theta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase theta subunit
U00001	trpI	Escherichia coli	tryptophan synthase iota subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase iota subunit
U00001	trpJ	Escherichia coli	tryptophan synthase jeta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase jeta subunit
U00001	trpK	Escherichia coli	tryptophan synthase kappa subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase kappa subunit
U00001	trpL	Escherichia coli	tryptophan synthase lambda subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase lambda subunit
U00001	trpM	Escherichia coli	tryptophan synthase mu subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase mu subunit
U00001	trpN	Escherichia coli	tryptophan synthase nu subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase nu subunit
U00001	trpO	Escherichia coli	tryptophan synthase xi subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase xi subunit
U00001	trpP	Escherichia coli	tryptophan synthase pi subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase pi subunit
U00001	trpQ	Escherichia coli	tryptophan synthase rho subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase rho subunit
U00001	trpR	Escherichia coli	tryptophan synthase sigma subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase sigma subunit
U00001	trpS	Escherichia coli	tryptophan synthase tau subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase tau subunit
U00001	trpT	Escherichia coli	tryptophan synthase theta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase theta subunit
U00001	trpU	Escherichia coli	tryptophan synthase iota subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase iota subunit
U00001	trpV	Escherichia coli	tryptophan synthase jeta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase jeta subunit
U00001	trpW	Escherichia coli	tryptophan synthase kappa subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase kappa subunit
U00001	trpX	Escherichia coli	tryptophan synthase lambda subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase lambda subunit
U00001	trpY	Escherichia coli	tryptophan synthase mu subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase mu subunit
U00001	trpZ	Escherichia coli	tryptophan synthase nu subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase nu subunit
U00001	trpA	Escherichia coli	tryptophan synthase alpha subunit	401	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase alpha subunit
U00001	trpB	Escherichia coli	tryptophan synthase beta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase beta subunit
U00001	trpC	Escherichia coli	tryptophan synthase gamma subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase gamma subunit
U00001	trpD	Escherichia coli	tryptophan synthase delta subunit	350	100.0	0.0	100.0	100%	100%	0	CD00001	tryptophan synthase delta subunit
U00001	trpE</											

[illegible]

Db 18617 GTGAGAGACCTGCACGAGAACTGTGAGAGTCTGGCGGAGCTGGCGCGACGCTCG 18676
 QY 658 GGGGTCGCGTCTTCTTCTTCCAGCGTGAAGGACCTGGCTGACACCAACCTGAGCGCC 717
 Db 18677 GAGGTGCGCTTCTTCTTCTTCCAGCGTGAAGGACCTGGCTGACACCAACCTGAGCGCC 18736
 QY 718 GCGTACTGTGATCCGCAACCTGACACGAGCGCGTCCGTTTCAGCATGCGCTGAGCGCC 777
 Db 18737 GCGTACTGTGATCCGCAACCTGACACGAGCGCGTCCGTTTCAGCATGCGCTGAGCGCC 18796
 QY 778 GCGGATGACGAGACCGCGCTCTCTGAGTACAGCCCGACCCGCTGCTCCCGCC 837
 Db 18797 CTGCGCGGAGTACCGCGGATCTGAGTACAGCTGCAACCGGCTGCTGATGAGCG 18856
 QY 838 ATCGAAGACACCAACCGAAGACACCGCGGAGAGTCAACCGGATGCGAGCTCCGCGC 897
 Db 18857 GTGCAAGAGGCGATTCACGAGGCGCGGCTCCGCGCGCGCGCGCGCGCGCGCGCG 18916
 QY 898 GCGGACACGACACCGCGCGCTTC 921
 Db 18917 GACCAAGGCGGACACCGCGCTTC 18940
 RESULT 10
 AAQ46806
 ID AAQ46806 standard; DNA; 29879 BP.
 AC AAQ46806;
 XX
 DT 22-DEC-1993 (first entry)
 XX
 DE eryA region of S. erythraea chromosome.
 XX
 KM Saccarapolyspora erythraea; eryA: biosynthesis; polyketide; module;
 KM erythromycin; condensation; elongation; acyl chain growth;
 KM gene replacement; ss.
 XX
 OS Saccarapolyspora erythraea.
 XX
 FH Key location/Qualifiers
 FT CDS 3..10199
 FT /*tag= a
 FT /note= "ORF 1"
 FT 10218..20921
 FT /*tag= b
 FT /note= "ORF 2"
 FT 20922..29879
 FT /*tag= c
 FT /note= "ORF 3"
 XX
 PN W09313663-A.
 XX
 PD 22-JUL-1993.
 XX
 PE 17-JAN-1992; 92MO-US00427.
 XX
 PR 17-JAN-1992; 92MO-US00427.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Donadio S; Katz L, McAlpine JB;
 XX
 DR WPI; 1993-242804/30.
 CC P-PSDB; AAR4430-32.
 PT Biosynthesis of specific polyketide analogues esp. erythromycin
 PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
 XX microorganisms
 PS Claim 27; Fig 2; 133pp; English.
 CC This sequence represents a fragment of the Saccarapolyspora erythraea
 CC genome, designated eryA. The polypeptides encoded by this region

CC are involved in the biosynthesis of the polyketide segment of
 CC erythromycin. eryA is organised in modules and each module takes care
 CC of one condensation step. The precise succession of elongation steps
 CC is dictated by the genetic order of the modules. This fragment may be
 CC specifically altered such that novel polyketide molecules of desired
 CC structure are produced. Three types of alteration may be produced;
 CC those inactivating a single function in a module which does not arrest
 CC acyl chain growth; those inactivating a single function in a module
 CC which does affect chain growth; and those affecting an entire module.
 CC The mutations may be introduced by gene replacement.
 XX
 SQ Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
 Query Match 41.8%; Score 385.4; DB 14; Length 29879;
 Best Local Similarity 65.6%; Pred. No. 1.7e-56;
 Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;
 QY 1 GTCTTGCTGTTTCCCGAGACGGGTGGGCAATGTTGGGATGGTGTGCTGCTGCC 60
 Db 11895 GTCTTGCTGTTTCCCGAGAGGTTTCGCAATGGGATGGTGTGCTGCTGCC 11954
 QY 61 TCCCTCCGCGTGTGCGCGCGGAGTGCAGAGGCGTCTGGCGCGCGTGGG 120
 Db 11955 TCGTCGCGGTGTGCGCGGAGAGTCCGGCTCGAGAGTCCAGTCCGAGTCCAG 12014
 QY 121 GACTGCTGTGTGAGACATCTGCGCGCGGAGCGCGGAGTCCGTTGGAGCGGCGC 180
 Db 12015 GACTGGAAGGTCTCCGAGGTGTGCGTACAGGCGCGCGCGCGCGCGCGTGA 12074
 QY 181 GATGTGTTCACGCTGTGTGCTTTCAGCGTCAATGTTCTTTGGCTGTGTGGCTTC 240
 Db 12075 GAGTGTGTACACCGCGTGTGTTCGCGGTGATGTGTGTGTGTGTGTGTGTGT 12134
 QY 241 TACGGATGAAACCGAGCGCGTCTTGGGATTCCTCCAGGAGAGATGCGCGCGCGAT 300
 Db 12135 TACGGGTGAGACCGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12194
 QY 301 GT 360
 Db 12195 GT 12254
 QY 361 CTGGCGCGT 420
 Db 12255 ATGCGGT 12314
 QY 421 GAGGAGCTCATGT 480
 Db 12315 CGGAGCGCTGT 12371
 QY 481 CGCTCACCGCGCTGT 540
 Db 12372 CGGT 12431
 QY 541 GCGACCGGCGGT 600
 Db 12432 GCGGAGGCGATCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12491
 QY 601 CAGCCCGT 660
 Db 12492 GAGCGGCTGT 12551
 QY 661 GTGCGGTTCTTCTTCCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
 Db 12552 GTGAGCTTCACTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12611
 QY 721 TACTGTATCCGGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
 Db 12612 TACTGTATCCGGAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12671
 QY 781 GATGAGGACACCGGT 840
 Db 12672 GAGTGGGATGAGAGCGCTTATGAGGTACGCGCGCATCGGTGTGTGTGTGTGTGT 12731

QY 841 GAAGACACCGACCGCCGGAAGACGTCACCGCGATCG 883
DB 12732 GAGGAGCGCGCTCGAAGAGCGTGAACGTCGCCGGAAGACCGCGCTCG 12774

RESULT 11
ID AAA09469 standard; DNA; 50937 BP.
AC AAA09469;
XX 29-AUG-2000 (first entry)
DT
XX Streptococcus oleandomycin gene cluster.
DE
XX
KM oleandomycin; oleandrolide; polyketide synthase; oleA; oleAII; oleAIII;
KM PKS; type I 8,8a-deoxyoleandrolide synthase; modular; ketosynthase;
KM acyl-transferase; acyl carrier protein; inactivated; polyketide;
KM macrolactone; antibiotic; molitide; erythromycin; 8a.
XX
OS Streptococcus antibioticus.

Key Location/Qualifiers
FT CDS 152..1426
FT FT /tag= a
FT FT /label= oleI
FT CDS complement (1528..2637)
FT FT /tag= b
FT FT /label= oleN2
FT FT complement (2658..4967)
FT FT /tag= c
FT FT /label= oleR
FT CDS 5772..18224
FT FT /tag= d
FT FT /label= ORF1
FT CDS /product= 8, 8a-deoxyoleandrolide_synthase_1
FT FT 18267..29717
FT FT /tag= e
FT FT /label= ORF2
FT CDS /product= 8, 8a-deoxyoleandrolide_synthase_2
FT FT 29787..40346
FT FT /tag= f
FT FT /label= ORF3
FT CDS /product= 8, 8a-deoxyoleandrolide_synthase_3
FT FT 40625..41830
FT FT /tag= g
FT FT /label= oleP1
FT CDS 41878..43158
FT FT /tag= h
FT FT /label= oleG1
FT CDS 43163..44443
FT FT /tag= i
FT FT /label= oleG2
FT CDS 44433..45173
FT FT /tag= j
FT FT /label= oleM1
FT CDS 45251..46411
FT FT /tag= k
FT FT /label= oleY
FT CDS 46491..47714
FT FT /tag= l
FT FT /label= oleP
FT CDS complement (47808..49517)
FT FT /tag= m
FT FT /label= oleB

XX
XX WO200026349-A2.
XX
XX 11-MAY-2000.
XX PD
XX 22-OCT-1999; 99MO-US24478.
XX PF
XX 29-OCT-1998; 98US-0106100.
XX PR

PR 16-FEB-1999; 99US-0120254.
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Betlach MC, Shah SK, McDaniel R, Tang L;
PI WPI; 2000-365602/31.
DR P-PSDB; AAY92707, AAY92708, AAY92709.
XX
XX Recombinant DNA compound encoding oleandrolide polyketide synthase for
PT synthesizing polyketides comprising a coding sequence for a domain of a
PT loading module or any one of extender modules
XX
XX Disclosure; Page 14-26; 86pp; English.

CC This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC The oleandrolide polyketide synthase (PKS), also known as
CC 8,8a-deoxyoleandrolide synthase, is encoded by three open reading frames
CC (ORF), designated oleA, oleAII and oleAIII. The PKS is a type I
CC "modular" enzyme, where each ORF encodes 2 extender modules and
CC the first ORF also encodes the loading module. Each module is composed
CC of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC acyl carrier protein (ACP) domain. The oleandrolide PKS loading module
CC contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC glutamine, present instead of the active site cysteine required for
CC activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC of polyketide macrolactones through multistep pathways involving
CC decarboxylative condensations between acylthioesters followed by cycles
CC of varying beta-carbon processing activities. The macrolide product of
CC the PKS, 8,8a-deoxyoleandrolide, is further modified by epoxidation and
CC glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC invention concerns an isolated recombinant DNA compound, comprising a
CC coding sequence for a domain of loading module or any one of extender
CC modules 1-4 or 1-6, including an oleandrolide PKS operably linked to a
CC promoter. Also discussed are recombinant oleandrolide PKS in which the
CC module 1 KS domain is inactivated by deletion or other mutation. In
CC particular, the inactivation is mediated by a change in the KS domain
CC that renders it incapable of binding substrate (the KS1-o mutation).
CC rendered by mutation in the codon for the active site cysteine. The
CC oleandrolide PKS is useful for synthesizing polyketides, which are useful
CC as antibiotics and molitides. Heterologous expression of oleandrolide PKS
CC in host cells such as Streptomyces coelicolor and S. lividans is also
CC made possible. Unmodified oleandrolide compounds can be provided to
CC cultures of Saccharopolyspora erythraea and converted to the
CC corresponding derivatives of erythromycins A-D.

XX
SQ Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;
Query Match 40.6%; Score 373.8; DB 21; Length 50937;
Best Local Similarity 64.4%; Pred. No. 1,5e-54;
Matches 593; Conservative 0; Mismatches 322; Indels 6; Gaps 2;

QY 1 GTCTTGCTGTTTCCGGACAGGCTGGGCAAGTGGGCGCGGAGTGGTGTCTCTGCC 60
DB 24645 GTGTTGCTGTTTCTGTCAGGCTTCGACAGTGGGAGTGGGCGCTGGTGTGAT 24704

QY 61 TCTCTCCGCTGTTTGGCGGCGATGACAGGCTGCGAGAGGCTTGGCGCGCTGGG 120
DB 24705 GCGTGTCCGCTGTTTGGCGAGCGGCTGGCGAGTGTGCCGCTGTGATCCGGTAGC 24764

QY 121 GACTGCTGTGTGTGACATCTTGGCGCGGAGCGCGGGGATGCGGTGGAGCGCGCC 180
DB 24765 GGTGTGCTGTGTGTGAGGTGTGGC---GGTGTGTGTGTGTGTGTGTGTGTGT 24821

QY 181 GATGTGTGTGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTC 240
DB 24822 GATGTGTGTGACCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 24881

QY 241 TACGGATGTGACCGCGAGCGGCTCTTGGCCATTCGCGAGGGAGATGCGCGCGCGCAT 300
DB 24882 TACGGATGTGACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 24941

QY 301 GTGTGTGGGCGCTGTGACCTGTGAGACCGCGCGAAGACTGTTCGCTGTGCGACCGCGGCG 360

Db 24942 GTGGCTGGGGGTTGAGTCTGGCCGATGTCGCGGGGTGATGTTGTCGAGACCGGGCG 25001
 QY 361 CTGGCCGCTGTGGCGGGCCGGGCGCATGAGCCTCAATGCGCGTGTCTGCGCAGAGATG 420
 Db 25002 ATGCGCCCGATCCCTGTGGGGGGGATGATGTCCTGCGTCTTTTACG---TGAGCGT 25058
 QY 421 GACACAGCTCATTTGTGAGCGGTGCGGGCGGTTGTGGGTGGCGGGCGTCAACGAGCC 480
 Db 25059 GTCCGACCATGCTCGACACCTACGATGTCGGGTTTCGGTCCGCGCGTCAATGGCCG 25118
 QY 481 CGTTCACCGCCGCTCTCGGGGATGCCGAGCGGTGACAGAGTGTGCGTACGTCGTC 540
 Db 25119 TCTTCGACCGCTCTCGTCGCGGTGACGTCACAGCCCTGGATGATGTTGTGGCGGTTGTGAG 25178
 QY 541 GGCACCGGGGTGGGGCGCGCGCGGTCGATGCCGTGACATATGCCCTGACATGCCCCCATG 600
 Db 25179 CGGAGGGGTGTCGGGCTCTGTCGTCCTCGGTGACTATGCTCCCACTCCGCGCAGATG 25238
 QY 601 CAGCCCTCGGGAGAGATTGTGAGCTGTGGGGGACATCAGCCCGCAGCCGTCGCGC 660
 Db 25239 GACCAATTAGCGATGAGCTGTGAGAGCGCTGGCGGACATCACTCCGCAACTTCCAGT 25298
 QY 661 GTGCCGTTCTTCTTCACGATGAGGGGCACTGCTGACACCAACCTTGAAGCGCGCC 720
 Db 25299 GTTCCGTTCTTCTGACGAGTGAAGCGGACCTGCTGACACGCGCTGATGCGGGG 25358
 QY 721 TACTGATACCGCAACCTGACACAGCGGTCCTTTTACGAGATCCGTCGAGCGCTGGCG 780
 Db 25359 TACTGATACCAATCTGCGGAGACGCTCGGTTCCAGAAACCGCTGAAAGGCTGTG 25418
 QY 781 GATGACGACACCGCGCTCTCTGTCGAAATGACGCCGCCACCCCTGTCGCCCGCATC 840
 Db 25419 GCTCAGGGGATGAGGCGGTCGTGTGAGTGAAGCCGACCCGCTCTGTCGCCGATC 25478
 QY 841 GAAGAACAACCCGAAAGACCGCCGAGAGAGTCAACGCGATCGGACGCTCGCGCGCGC 900
 Db 25479 GACGACGACCTCTGACGCGCTCGACAGAGAGCGCGCTACTCGGCTCCCTGCGCGTGAAC 25538
 QY 901 GACACGACACCGCGCGCTTC 921
 Db 25539 GAAAGCGGCTGACGCGACTC 25559

RESULT 12

AAV99255 standard; DNA; 1010 BP.

AAV99255;

09-MAR-1999 (first entry)

DNA encoding an active acyltransferase domain (NidA75) from a PKS.

Enzymatically active acyltransferase domain; polyketide synthase; PKS;

polyketide synthesis; antimicrobial; anticancer; antifungal;

immunosuppressant; antihelminthic; NidA75; ethyl acyltransferase domain;

module 5; PKS cluster; de.

Streptomyces caelestis.

MO9851695-A2.

19-NOV-1998.

13-MAY-1998; 98MO-US09518.

16-MAY-1997; 97US-0858003.

(ABBO) ABBOTT LAB.

Kakavas SJ, Katz L, Pereda-Lopez A, Ruan X, Staasi DL;

Summers RG;

XX WPI; 1999-070114/06.
 DR P-PSDB; AAW87714.
 PT New polyketides produced by microorganism having altered polyketide
 PT synthase genes - especially chimeras containing a heterologous
 PT acyltransferase domain, useful as e.g. antimicrobials, anticancer
 agents
 PS Claim 24; Fig 21; 137pp; English.
 CC The present sequence encodes an enzymatically active acyltransferase
 CC domain, NidA75, from a polyketide synthase (PKS). The NidA75 protein
 CC is derived from the ethyl acyltransferase domain from module 5 of the
 CC PKS cluster of Streptomyces caelestis NRRL-2821. The sequence
 CC can be used to replace acyltransferase domains of other PKS enzymes,
 CC which are then used to generate novel polyketides by de novo biosynthesis
 CC rather than by chemical modification. Polyketides of the invention have
 CC the methyl groups on the macrolactone ring substituted with -H, -Et
 CC and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
 CC dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
 CC antimicrobial, anticancer, antifungal, immunosuppressant and/or
 CC antihelminthic agents, or their intermediates.
 SQ Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;
 Query Match 39.4%; Score 362.8; DB 20; Length 1010;
 Best Local Similarity 63.7%; Pred. No. 1.3e-52;
 Matches 585; Conservative 0; Mismatches 327; Indels 6; Gaps 2;
 QY 1 GTCTTGTGTTTCCGGAACAGGTTGAGGACAGTGGGCGGAGATGATGTCGTCGCGC 60
 Db 13 GTTTCGATGTTCCCGGCAAGGCTCGCAGTGGCGGATGCGGAGGGGCTGCTGAG 72
 QY 61 TCTCTCCGATTTGCCCGCGGATGACAGCGCTGCGAGAGCTTCGCGCGCTGGGGT 120
 Db 73 CGGTCCGCGGCTTCGAGATGTCGCGCGGACGCTGTGCAAGCGCGCTCGCGCTGACCTC 132
 QY 121 GACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 Db 133 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
 QY 181 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 Db 193 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
 QY 241 TACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 Db 253 CTGGGGGTGGAACCGCGCGGCTCTCGGCGCACTGCAAGGGATGATGATGATGATGAT 312
 QY 301 GTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 Db 313 GTGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
 QY 361 CTGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 Db 373 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
 QY 421 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
 Db 433 CGGCGCGGCT---GATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 489
 QY 481 CGCTCACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 Db 490 GGTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
 QY 541 GGCACCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 597
 Db 550 GGTGAGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609
 QY 598 GTGACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 657
 Db 610 GTGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669

Db 910 CGCCACGCGCTCTCGC 927

RESULT 14

AAV21187

ID AAV21187 standard; DNA; 53789 BP.

XX

AC AAV21187;

XX

DT 24-JUL-1998 (first entry)

XX

DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.

XX

KM Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;

XX

polyketide synthase; actinomycete; ansamycin; ds.

OS Amycolatopsis mediterranei.

XX

XX Key Location/Qualifiers

FT CDS 1825..15543

FT /tag= a

FT /label= ORF_A

FT /product= "polyketide synthase"

FT 15550..30759

FT /tag= b

FT /label= ORF_B

FT /product= "polyketide synthase"

FT 30895..36060

FT /tag= c

FT /label= ORF_C

FT /product= "polyketide synthase"

FT 36259..41325

FT /tag= d

FT /label= ORF_D

FT /product= "polyketide synthase"

FT 41373..51614

FT /tag= e

FT /label= ORF_E

FT /product= "polyketide synthase"

FT 51713..5293

FT /tag= f

FT /label= ORF_F

FT /product= "polyketide synthase"

PN MO9807868-A1.

XX

PD 26-FEB-1998.

XX

PF 18-AUG-1997; 97MO-BP04495.

XX

PR 20-AUG-1996; 96BP-0810551.

XX

(NOVS) NOVARTIS AG.

XX

PA Engel N, Schupp T, Toupet C;

XX

PI WPI: 1998-169172/15.

XX

DR P-PSDB; AAW52845-W52850.

XX

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used

XX

PS to produce rifamycin and rifamycin analogues

XX

Claim 4; Page 53-102; 205pp; English.

XX

XX The present sequence represents a Amycolatopsis mediterranei rifamycin

CC synthesis gene cluster DNA fragment from the present invention. The

CC DNA fragment comprises a DNA region involved directly or indirectly

CC in the gene cluster responsible for rifamycin synthesis, including

CC the adjacent DNA regions to the right and left which, by reason of

CC their function in connection with rifamycin biosynthesis, qualify

CC as constituents of this rifamycin gene cluster, and functional

CC fragments, derivatives or constituents of these. The Amycolatopsis

CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used

CC for producing rifamycin, rifamycin analogues or precursors. It can also

CC be used for inactivating or modifying genes involved in ansamycin or

CC rifamycin biosynthesis. The DNA can be used for constructing mutant

CC actinomycetes strains from which the natural rifamycin or ansamycin

CC biosynthesis gene cluster has been partly or completely deleted. The

CC DNA fragment can be used for assembling a library of polyketide

CC synthases, which can be used for assembling a library of polyketides.

CC A hybridisation probe of the invention can be used for identifying DNA

CC fragments involved in the biosynthesis of ansamycins.

XX

SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

XX

Query Match 38.1%; Score 351; DB 19; Length 53789;

Best Local Similarity 64.8%; Pred. No. 1e-50;

Matches 589; Conservative 0; Mismatches 305; Indels 15; Gaps 4;

QY 1 GTCCTCGTGTTCCTCCGACAGGCTGAGTGGCCCGGAGTGGTGTCTGCTCCG 60

DB 17257 GTCCTGAGTGTTCCTCCGACAGGCTGAGTGGCCCGGAGTGGTGTCTG 17316

QY 61 TCCCTCCGAGTGTTCCTCCGACAGGCTGAGTGGCCCGGAGTGGTGTCTG 120

DB 17317 TCCCTGAGTGTTCCTCCGACAGGCTGAGTGGCCCGGAGTGGTGTCTG 17376

QY 121 GACTGCTGTGTGATGACATCTCTGCGCGGAGACGCGGGGATGCCGTGTGAGCGCGC 180

DB 17377 GACTGCTGTGATGACATCTCTGCGCGGAGACGCGGGGATGCCGTGTGAGCGCGC 17430

QY 181 GATGTGTTCAGCTCTGTGCTGTTACAGCTCATGTGTCTTTGGCTGTCTGTGCGCTTC 240

DB 17431 GATGTGTTCAGCTCTGTGCTGTTACAGCTCATGTGTCTTTGGCTGTCTGTGCGCTTC 17490

QY 241 TACGGTATGAAACCCGAGCGCGGCTCTTGGCCATTCACAGGGAGATGCGCGCGCAT 300

DB 17491 TACGGTATGAAACCCGAGCGCGGCTCTTGGCCATTCACAGGGAGATGCGCGCGCAT 17550

QY 301 GTGTGTGGGCGCTGAGCTGAGAGACGCGCGGAGACATGTTGCCCTGCGACCGCGGCG 360

DB 17551 GTGTGTGGGCGCTGAGCTGAGAGACGCGCGGAGACATGTTGCCCTGCGACCGCGGCG 17610

QY 361 CTGGCCGCTGTGCGG---GGCCGGGCGGATGCTCATGTGCTGCTGCTGCTGCTGCTG 417

DB 17611 ATGCGCGCTCTGCTGCGGCGGCGGCGGCGGATGCTCATGTGCTGCTGCTGCTGCTG 17670

QY 418 GTGAGACGCTCATGTGAGAGCGGTGGGCGGGGCTGTGGGTGGCGGCGGCTCAACGCG 477

DB 17671 GTGAGACGCTCATGTGAGAGCGGTGGGCGGGGCTGTGGGTGGCGGCGGCTCAACGCG 17727

QY 478 CCCGCTCCACCGCGCTCTGCGGGGATGCCAGGCGGTGACGAGTGTGCTACTGT 537

DB 17728 CCCGCTCCACCGCGCTCTGCGGGGATGCCAGGCGGTGACGAGTGTGCTACTGT 17787

QY 538 GCGGACCGGGGCTGCGGCGCGGATCCCGTGCATGTGCTGCTGCTGCTGCTGCTGCTG 597

DB 17788 GCGGACCGGGGCTGCGGCGCGGATCCCGTGCATGTGCTGCTGCTGCTGCTGCTGCTG 17847

QY 598 GTGACGCGCTGCGGAGAGTGTGAGAGCTGTGGGGGACATACGCGCGAGCGCTTC 657

DB 17848 GTGACGCGCTGCGGAGAGTGTGAGAGCTGTGGGGGACATACGCGCGAGCGCTTC 17907

QY 658 GCGGTGCGCTTCTTCTCAACGCTGAGGAGCACTTGCT---GACACCAACCACTTGAGC 714

DB 17908 GCGGTGCGCTTCTTCTCAACGCTGAGGAGCACTTGCT---GACACCAACCACTTGAGC 17967

QY 715 GCGCTTACTGTGATCCGCACTGACACAGCGGCTGCTTTCAGGATGCGCTCCAGGCG 774

DB 17968 GCGCTTACTGTGATCCGCACTGACACAGCGGCTGCTTTCAGGATGCGCTCCAGGCG 18027

QY 775 CTGGGAGTGAAGGACACCGCTCTTGTGCAAGTCAACCCCAACCCCACTGCTGCC 834

DB 18028 CTGGGAGTGAAGGACACCGCTCTTGTGCAAGTCAACCCCAACCCCACTGCTGCC 18087

QY 835 GCATCGAAGACACCAAGAGACACCGCCGAAAGCTACCGCGATGCGGACGCTCGCG 894

THIS PAGE BLANK (USPTO)

Qy	1258	ATCGCGGCACGAATCGCATGTGATTTTGGAGGCGCGCG	1299
Db	6586	ATCAGTGGCACCACCGGCACCTGATCCTTGAACACCTCCG	6627

Search completed: June 17, 2003, 13:48:14
Job time : 269.094 secs

RESULT 15

AAF88338 standard; DNA; 14775 BP.

AAF88338;

28-AUG-2001 (first entry)

S. spinosa DNA fragment encoding ORF21, SEQ ID 47.

Forosamine; trimethylrharnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.

Saccharopolyspora spinosa.

DE19957268-A1.

08-MAR-2001.

29-NOV-1999; 99DE-1057268.

27-AUG-1999; 99DE-1040596.

(FARB) BAYER AG.

Eberz G, Moehrl V, Froede R, Velten R, Salas JA;

WPI; 2001-267102/28.

P-PSDB; AAB70968.

New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
recombinant production of insecticidal spinosyns and their derivatives

Claim 7a; Page 239-264; 354pp; German.

This invention describes a novel method nucleic acid (I) and its encoded
polypeptide (II) containing at least one region that encodes an enzymatic
activity involved in biosynthesis of spinosyns. (I) are used (i) to
identify, inactivate or modulate genes involved in the biosynthesis of
(II); (ii) to generate a library of polyketide synthases; (iii) for
adding fotsamine or trimethylrharnose to a spinosyn or polyketide
aglycone; and (iv) for recombinant production of the corresponding
enzymes, which are used for production of (II), their precursors or
derivatives, including production of transgenic plants that express (II)
and thus have increased resistance to insects. (I) are also useful as
markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
macrolides with insecticidal, but not antibacterial, activity, and can
also be used to raise specific antibodies, useful for identifying
expression clones in a gene bank. Cells transformed with (I) may produce
(II) at significantly increased levels or produce new derivatives of
(II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase.

Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 other;

Query Match

Best Local Similarity 50.0%; Score 650.4; DB 22; Length 14775;

Mismatches 918; Conservative 0; Mismatches 366; Indels 18; Gaps 3;

QY 1 GACCGATCGCCATCATCGGATGCGATCGCTTTCCGGGGGAGTCCGGTCCCGGAAG 60
DB 5341 GATCTGTTCGATTGTCCGGATGGATGTCGTTTCCGGGTGGGGTGTCTGCGCGGAG 5400
QY 61 GACCTGTGGAGCTCCGCGCTTCGGGCGGAGACGCCATCGGCGCTTCCCGACCGCCG 120
DB 5401 GAGCTTTGGCGCTGGTGGCGGGCGTGGATGCGGTGGCTGCCAGACGATCGC 5460
QY 121 GGATGGCCACCGAACAGCGTACGCCAGGACCCACGAGCCCGGACAGTTCTATCCG 180
DB 5461 GGCTGGGATCTCGCGGGCTGTGTACGATCCTGATCCCGATCGTCTCGGACCTCGTATGTG 5520

QY 181 CAGGAGGCGGGTTCCTTTCAGACGCGCGCATCTTCAGCGCGGCTTCTTCGGAATCACT 240
DB 5521 TGTGAGGCGGGTTCCTGCGGACGCGCGGAGTTTCGATGTCATGTTTCGCGATCAGC 5580
QY 241 CCACTGAGGACATCGCGATGATCCGACGACGCGGCTGCTCTCGAGACGCTCTCGGAG 300
DB 5581 CCGCGTGGGGCTTGGCGATGATCCGACGACGCGGTTCTGCTCGAGGTCGCTCGGAA 5640
QY 301 GCGTTCGAGCGGGCGGAATCGATCCGCTGTCGTAACGCGGTCCTGACGCGGCTTC 360
DB 5641 ACCTTGGAGCGGCTGGGATCGATCCGTTCTGTTGACGCGGACCGGACCGGTGTGTT 5700
QY 361 CGCGCGCGCTCTCTCTTCCACTACCGCGCGCGGTATGACACCGCTGCTCGAGGCGCGC 420
DB 5701 CGCGGCTTGTATGATACACGACTATGGGGCCGATTCATTACAGAGCACCGAGGCTTC 5760
QY 421 GCGGACGTGGAGGGCCACATCTCTACCGGTACACGCGGAGCGTCTGTCGGGCGTATC 480
DB 5761 -----GAAGGGCACTCGGACGCGCAATGCGGGAGCGTGTCTGCGGTCCGGTT 5811
QY 481 GCCTACAGCTTCGGGCTGGAAGGCGCGGATCACCGTGGACACGGGTGCTCGGCATCG 540
DB 5812 CGGTATTCGTTGGTTTCGAGGGTCTCGGGTGACGGTGATACGGCGTGTCTGCTGCTG 5871
QY 541 CTGCTGACGCTGCACTCGCGTGCAGTCCGCTCGGTCGGGTGAGTGACACGCTCGCGCTG 600
DB 5872 TTGTTGGGTTTACACTGCGGGTCAAGCACTGCGGGCGGTGAGTGCGAATTCGCCCTT 5931
QY 601 GCGGCGCGCTCTCGGTCACTCCACCTCGGCACTGTTTCATCGATTTCTCCGGGAGCGC 660
DB 5932 GCGGTGGCGTCACTGATGTCGACGCGGACGAGTTCTGCGGAGTTCTCCCGTCAACGG 5991
QY 661 GGGTGTGCGTGGAGCGGAGTGCAGCGGTACTCGGCTGCGACGCGGACCGGCTGG 720
DB 5992 GGTCTGGCTCCGGATGGCGGTGCAAGTCTGTCGGCGCGCGCGGATGGACACCGGGTGG 6051
QY 721 GCGGAGGCGCTCGGGATGCTGTGTTGGAGCGGTGTCGGATGCGGTGCGGTGGGGCAT 780
DB 6052 GCGGAGGCTCGGGTCTGTTGTTGCTGAGCGGTGTCGGATGCGCGGGCAATGGGCAC 6111
QY 781 CGGTTGTCGGGTGTTGATCGCGGAGTCCGTTCAACAGGACGCTGCTCGAATGGGCTG 840
DB 6112 GAGGTTCTGGCGGTGTTGCGGGTAGCGCGGTGAACAGGACGCGCGTCTGAAATGGCTG 6171
QY 841 ACGGCGCGACGCTCGGCTCAGGAGCGGTTGATCCGGCAGGCTTGGCGAACCGCGGG 900
DB 6172 ACTGCGCAATGGTCCGTCACGAAAGGGTGTATCACCCAGGCACTCACGAGTCCCGGG 6231
QY 901 TTGTCGTCGGCGGATGTCGATGTTGGAGGGGACGCGGACGCGACGCTGGGTGAT 960
DB 6232 CTGTCGCTTCGACGTTGGATGCTGTGGAGGCGCATGGGACGCGGACGCGGCTTGGTAT 6291
QY 961 CCGATCGAGGACAGCGGCTGCTCCGACGCTACG---GGCAGCGGGCGGTTGACAGCGCG 1017
DB 6292 CCGATCGAGGCGCAGGCGTTCGATCGTACGTCAGCGCGGGATCGTGATCCCGGTGCGCG 6351
QY 1018 CTGTGGCTGGGGTCTCTCAAGTCCAACTCGGCGACACCATGGCTGCGCGGGTGGGT 1077
DB 6352 TTGTGGCTGGGGTTCGGTGAAGTTCGAATATTGTTCAACCCAGGCGCGCGGTTCGCT 6411
QY 1078 GGGGTCTCAAGATGTTGATGGCGTTGGGAGGGGTGTTGCGCGGACGTTTGCATGTG 1137
DB 6412 GGTGTGATCAAGATGTTGATGCGCATGCGGACGAGGGAGTGTCGCGGACGCTTGCACGTG 6471
QY 1138 GATAAGCGCTCGCGCAGGTTGACTGTGTCGCGGGGCGGTGCGGCTCTGACGAGGCG 1197
DB 6472 GACGAGCCCTCCGCGCAGTGGTCTGCTGCGGCGACGGTCCAACTCCTCACGAGAAC 6531
QY 1198 GTGCGGTGGCGGGGACCGGACGAGGCGGTGCGGCGGGCGGAGTGTCTGTTCTGGG 1257
DB 6532 ACGCCCTGGCC-----CGACAGCGGTGCTTTCGCGGGCGGGCGTGTCTATGTTGGG 6585

XX (STAM) DSM NV.
 XX PA Martin JP, Aparicio JF, Colina AJ;
 XX PI WPI; 2001-080693/09.
 XX DR P-PSDB; AAB31557.
 XX
 PT New polynucleotides encoding enzymes involved in the biosynthesis of
 PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
 PT the synthesis of new compounds -
 XX
 PS Disclosure; Page 39-46; 116pp; English.
 XX
 CC The present sequence encodes a polyketide synthase which is associated
 CC with the biosynthesis of pimaricin. The polyketide synthase polypeptide
 CC is useful for the oxidative modification of a methyl group of a suitable
 CC compound, e.g. a bioactive compound including a secondary metabolite,
 CC antibiotics and anticancer agents. Recombinant cells comprising the
 CC gene are useful for the production of pimaricin. The polyketide synthase
 CC polynucleotide may be over expressed in Streptomyces, leading to an
 CC increase in the biosynthesis of pimaricin, as a source of primers for
 CC amplification reaction and as probes.
 XX
 SQ Sequence 5544 BP; 699 A; 2052 C; 2023 G; 770 T; 0 other;
 Query Match 50.9%; Score 662.6; DB 22; Length 5544;
 Best Local Similarity 71.1%; Pred. No. 6.2e-105;
 Matches 926; Conservative 0; Mismatches 359; Indels 18; Gaps 3;
 1 GACCCGATCCCATCATCGCATGCGATGCGGTCCTCCCGGGGAGTCCGGTCCCGAAG 60
 1981 GACCCGATCGCATCGTGGGATGCTCGCGCTCCCGGGGAGCGACTCCCGGAC 2040
 61 GACCTGTGGAGTGGCCGCTCGGGCGGAGAGCGCATCGGGCGTTCCTCCCAACGACCGC 120
 2041 GCCCTGTGGAGTGTCTCCCGACGCGGACCGAGCGGATGTCCCTTCCCAACGACCGC 2100
 121 GGATGCCCCACAGGACAGCGTCACGCCAGGACCCACGACCGCGGACGTTCTATCG 180
 2101 GGCTGGGACTGGACCGGCTGTTCGACGAGGATGCGGACCGCGGGGTACCTGTACGCC 2160
 181 CAGGAGGCGGTTCTTCACAGCGCGGCACTTCGACCGCGGCTTCTTCGGAATCACT 240
 2161 CGCGAAGCGGCTTCTCTCAAGCGCGCGGCTTCGACCGCGGCTTCTTCGCGCTGCG 2220
 241 CCACGTAGGCACTGGCGATGATCGGAGAGCGGCTCTCTGTGAGAGCGTCTGGAG 300
 2221 GACCAAGGAGCGACGCGGACCGATCCGACGAGCGGCTGCTTCGGAGCGGCTGGGAG 2280
 301 GGGTTCGAGCGGCGGGAATCGATCGCTGTCGGTACGGGCTCCGTACGGGCGTCTTC 360
 2281 ACCTTCGAGCGGCGGCACTGACCGGCTGACCGGAGTCCCTGAGGGGAGCGGTACGGGCGTTC 2340
 361 GGGGCGGCTCTCTTCGACTACGCGCGCGATGGAACACGCGCTGCTGTGAGGCGCGC 420
 2341 ACGGGCGGATGACCGGCGCTATGGAACACGCGGCTCGCGCGCGCGGCGCATG---- 2396
 421 GCGGAGTGGAGGCGACATCTCTACCGGTAACAGGGGAGCGTCTGTGCGGCGGCTATC 480
 2397 -----GGAGAGATGCTCATCACCGGACCGCGGCGAGCGGCTCTCGGGGCGCATC 2448
 481 GCCTACAGCTTCGGGCTGAAGCGCGGCGATCACGCTGGACACGGGCTCTCGGCATCG 540
 2449 GCCTACACCTACGGGCTGAAGCGCGCGGCTGACGCTGACACCGCTCTCTGCTCTCC 2508
 541 CTCGTACGCTCATCTGGCGTGCACGCTGCGGCTGGGTGAGTGACGCTCGCGGTG 600
 2509 CTCGTGCGGCTCATCTGGGCTGCGGCTGCGCTCGGCGGAGACCGACCTGCGGTG 2568
 601 GCGGCGGCGTCTCGGTATGTCACCTCTGGGATGTTTCATCGATGTTCTCCCGGAGCGC 660
 2569 GCGGCGGCGTCACTGCGGACCGCGGCGGCTTCGACACTTCTCTCCCGGTGCGC 2628

QY 661 GGGCTGTCTGGTGGACGGCAGGTGCAAGGCGTACTCGGTGCAAGCGACGCGACCGGCTCG 720
 DB 2629 GGGCTGTCTCCCGACTCCCGCTCCATGGGTACGCGGACCGCGCGGAAACGGCTCGGCTCG 2688
 QY 721 GCGGAGGCGTCTGGATGCTGTGTGTGAGCGGTTGTGCGATGCGGTCGCGTGGGGCAT 780
 DB 2689 TCGGAGGCGCGGGGCTGCTTCTGTGAGCGGCTGAGCGACCGCGCGGCAACGACAC 2748
 QY 781 CGGCTGTCTGGGCTGTGACGCGGAGTGGGTCAACGAGGCGGTGCGTTCGATGGGCTG 840
 DB 2749 CGTGTCTTGGGCTGCTGTACGGGCTCCCGCGTGAATCAGGACGGGCGCTTCCAAACGGGCTC 2808
 QY 841 ACGGCGCGCAACCGGTCTCGGCTCAGGAGCGGGTGATCCGCGAGCGGTGTGGCGAACCGGGG 900
 DB 2809 ACCGCGCGGAGCGGACCGCGACAGAGCGGTCTATCCGCGAGGCTTGGCGCGCGCGG 2868
 QY 901 TTGTCCGTGGCGGATGTGTGTGTGAGGGGCGACGGGACCGGACGACGCTGGGTGAT 960
 DB 2869 CTGACGCGGAGGACGTGGAGCGCTGTGAGGGGCGACCGGACCGCGCTCGGCGAC 2928
 QY 961 CCGATCGAGGCGACAGGCGGTGCTCGCCACGTTACGG---GCAGGGGCGGTCGACAGGCG 1017
 DB 2929 CCGATCGAGGCGAGGCGCTGCTGGCCACGTACGCGGCAACAGCGGCTGTGGAAACGGCG 2988
 QY 1018 CTGTGGCTGGGCTCTCTGAAGTCCAAACATCGGGGACACATGGCTGCGCGGCTGTGGGT 1077
 DB 2989 TTGTGGCTGGGCTGGTGTGAGTCTGAGTCTGGGACACACAGCGCGCGCGGCTGCTC 3048
 QY 1078 GGGGTCTATCAAGATGTGTGTGTGCGGGAGGGGTGTTGCCCGGACGTTGATG 1137
 DB 3049 GCGGTCTCAAGACGGTCTCGCGTGGCGCACGCGCTCTGCGCGACGACGCTGACGCTG 3108
 QY 1138 GATAAGCGCTGCGCGCAGTGTGCTCGCGGGGGGGTCCGCTGCTGACGAGCGG 1197
 DB 3109 GACGCTCTCGGCGCAAGGTGCTGCTCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3168
 QY 1198 GTGCGCTGGCGGGGACCGCGAGCGGCGGTGTCGCGGGGGGGGAGTGTCTGCTGCTG 1257
 DB 3169 CGGCGCTGGCGACAGGGA---GAGCGGACGTACGCGCGCGGGGGGTCTCTGCTGCTG 1257
 QY 1258 ATCGCGCGACGAATGCGCATGTGATTTTGGAGGAGGCGCGCG 1300
 DB 3226 CTCACCGGACCAACGCGCACGTCGATCTCTGAGGAGGCGCGCG 3268
 RESULT 14
 AAV21187
 ID AAV21187 standard; DNA; 53789 BP.
 XX AC AAV21187;
 XX DT 24-JUL-1998 (first entry)
 XX DE Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
 XX KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 XX OS polyketide synthase; actinomycete; ansamycin; ds.
 XX PH Amycolatopsis mediterranei.
 XX FT Key Location/Qualifiers
 FT CDS 1825..15543
 FT /tag= a
 FT /label= ORF_A
 FT /product= "polyketide synthase"
 FT 15550..30759
 FT /tag= b
 FT /label= ORF_B
 FT /product= "polyketide synthase"
 FT 30895..36060
 FT /tag= c
 FT /label= ORF_C

PR 22-FEB-1996; 96US-0012078.
XX (ELIL) LILLY & CO ELI.
XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
XX WPI; 1997-418046/39.
DR P-PSDB; AAW22601-W22605.
XX
XX DNA encoding Streptomyces fradiae tylosone synthase domain - for
PT production of tylosin-related polyketide compounds
XX
XX Claim 2; Pages 8-66; 220pp; English.
XX This sequence represents the tylosone synthase gene cluster of the
CC invention. This sequence is also referred to as the tyIG gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the biosynthesis of the polyketide tylosone, isolated
CC from Streptomyces fradiae. Tylosone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 52.0%; Score 676.4; DB 18; Length 43280;
Best Local Similarity 72.0%; Pred. No. 2.4e-107; Indels 21; Gaps 3;
Matches 932; Conservative 0; Mismatches 341;
1 GACCCGATCGCCATCATCGCATGGCATGCGTTTCCCGGCGGAGTCCGGTCCCGAAG 60
31337 GAGCCGATCGCCGTCGTGGCATGGCTCGCGTACCCCGGAGCGGAACCCCGAA 31396
61 GACCTGTGGAGTGGCGCTCGGGCGGAGAGCCATCGGGCGTTTCCCGACCGACCGC 120
31397 CAGTTCTGGGAACTGTGTGACACCGGACCGGCGCGCGCGCGTCCCGACCGG 31456
121 GGATGCCACGGAACAGCGTACGCGCCAGGACCCCGGCGGCGGCGGCGGCGGCGG 180
31457 GGATGGGACACCGACCGGCTGTAGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGG 31516
181 CAGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCGCGGCTTCTTCGGAATCA 240
31517 CGGGAGGCGGCTTCCTTCACGACGCGGCGCACTTCGACGCGGCACTTCTTCGGAATTC 31576
241 CACGCTGAGCACTGGCGATGGATTCGCGAGCAGCGCTGTGTGAGAGAGCTCTCGGAG 300
31577 CCGCGTGAAGCGGTGGCGATGGATTCGCGAGCAGCGCTGTGTGAGAGAGCTCTCGGAG 31636
301 GCGTTCGAGCGGCGGGAATCGATCCGCTGCGTACGCGGTCGCCGTCGCGGCGGCGG 360
31637 GCGATCGAAGCGGCGGCTATCGACCGCGAGGAGCTCCGCGGCGAGCGCGGCGGCTAC 31696
361 CGCGGCGGCGCTTCCTTCGACTACGCGCGCGGTATGGAACACCGCGCTGTGTGAGGCGGCG 420
31697 GTGGGCGCTTGGGACAGCGCTACACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 31744
421 CGGACGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
31745 GCGGAGTTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31804
481 GCCTACAGCTTCGGGCTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
31805 GCCTACAGCTTCGGGCTGGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31864
541 CTCGTACGCTGATCTGCGTGCCTGAGTGTGCGGTTCGGGTGAGTGCAGCTTCGCGGCTG 600
31865 CTGGTCCGCTGCACAAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31924
601 GCGCGGCGGCGTCTCGGTATGTACCTCGGCGGATGTTTCATCGAGTTCTCCCGGCGGCGG 660
31925 GCGCGGCGGCGTGTGAGCGGTGATGGCGGACCCCGGCGGCTGTGTGTCAGTTTCGCGGCGGCGG 31984

QY 661 GGGCTGTCTGGTGACGCGAGCGTGAAGCGCTACTCGGCTGACGCGGCGGCGGCGGCTGG 720
DB 31985 GGGCTGGCGCGGAGCGGCGGCTGCAAGCGCTTGCAGCGCGCGCGCGGCTTCGCGCCC 32044
QY 721 GCGAGGCGCTCGGAGATGCTGTTGTGAGCGGTTGTGCGATGCGGTGCGGCTGGGGAT 780
DB 32045 GCGAGGCGTGGGAGATGCTGCTGTGAGCGGTTGTGCGATGCGCGCGGTTGGGGAT 32104
QY 781 CGGCTGCTGGCGGTGTACGCGGCGAGTCCGCTCAACAGGACGCTGCGTCAATGCGGCTG 840
DB 32105 CGGCTGTTGGCGGTGTGTGCGGTGCGCGGTGATCAGGACGCTGCGTCAATGCGTGTG 32164
QY 841 ACAGCGCGGCGGAGCGGCTCGGCTCAGAGCGGCTGATCCGCGGCGGCTTGGCGAAGCGGCG 900
DB 32165 ACAGCGCGGAGTGGTCCGTCAGAGCGGCTGATTCGTCAGGCGTGGGGAATCGCGCG 32224
QY 901 TTGTCCGTGGCGGATGTGATGTGTGAGGGGCAAGGAGCGGCGGCGGCGGCGGCGGCGG 960
DB 32225 TTGACGCTGGCGGATGTGATGTGTGAGGCGGATGCGGAGCGGCGGCGGCGGCGGCGG 32284
QY 961 CCGATCGAGGCGGCGGCGGCTGCTCGCCACGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017
DB 32285 CCGATCGAGGCGGCGGCGGCTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32344
QY 1018 CTGTGGCTGGGCTCTCTCAAGTCCAAACATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1077
DB 32345 GTGTGGTGGGCTGCTCAAGTCCGATGATGATGATGATGATGATGATGATGATGATGATG 32404
QY 1078 GGGGTCATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1137
DB 32405 GGTGTGATCAAGTGTGTGCGATGCGGTATGCGGTGATGCGGTGATGCGGTGATGCGGT 32464
QY 1138 GATAGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1197
DB 32465 GATAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32524
QY 1198 GTGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1257
DB 32525 CGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 32578
QY 1258 ATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291
DB 32579 GTGAGTGGTACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32612
RESULT 13
AAF24891
ID AAF24891 standard; DNA; 5544 BP.
XX
AC AAF24891;
XX
DT 20-APR-2001 (first entry)
XX
DE Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW anticancer; pimaricin; ss.
OS Streptomyces natalensis.
XX
FH Key Location/Qualifiers
CDS 1..5544
FT /*tag= a
FT /product= "polyketide synthase"
XX
PN WO200077222-A1.
XX
XX 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-EP06227.
XX
PR 14-JUN-1999; 99EP-0201893.

CC particular, the inactivation is mediated by a change in the KS domain
 CC that renders it incapable of binding substrate (the KSI-o mutation),
 CC rendered by mutation in the codon for the active site cysteine. The
 CC oleandolide PKS is useful for synthesizing polyketides, which are useful
 CC as antibiotics and motilides. Heterologous expression of oleandolide PKS
 CC in host cells such as *Streptomyces coelicolor* and *S. lividans* is also
 CC made possible. Unmodified oleandolide compounds can be provided to
 CC cultures of *Saccharopolyspora erythraea* and converted to the
 CC corresponding derivatives of erythromycins A-D.
 XX
 SQ Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

Query Match 52.9%; Score 689; DB 21; Length 50937;
 Best Local Similarity 72.3%; Pred. No. 1.7e-109;
 Matches 944; Conservative 0; Mismatches 340; Indels 21; Gaps 3;

QY 1 GACCCGATCGCATCATCGGCATCGCATCGGCTTCCCGGGCGGAGTCGGTCCCGGAG 60
 DB 13407 GAGCCGATCGCATCGTGGGAATGGCATCCGCTTCCCGGGAGGCATCGCTTCCCGGGAG 13466

QY 61 GACCTGTGGAGTGGCCGCTCGGGCGGAGAGCGCATCGGCGGTTCCCCACCGACCGC 120
 DB 13467 GATCTGTGGAGTGGCTACCGTGTGTAGAGCTGTGTCTCCGACTGCGGAGACCGC 13526

QY 121 GGATGCCCCACGGAACAGCGTCAAGCCAGGACCCCAACGACGCCGCGGACGTTCTATCCG 180
 DB 13527 CGGTGGGACTTGGGGCGCTGTACGATCCGATCCGAGCCGCGCGGTACGTACATATG 13586

QY 181 CAGGAGGGCGGTTCCTTACGACGCGGCGCATTCGACGCGGCTTCTTCGGAATCAGT 240
 DB 13587 CGTAGGGGTGCTTTCCTGACAGGCGGGGAGTTTCGACGCGGCTTCTTCGGGTATCTCG 13646

QY 241 CCACGTGAGCGACTGCGCATGATCCGACGAGCGCTCTCTGTGAGAGCGTCTTGGGAG 300
 DB 13647 CCGGTGAGCGGTGCGCATGAGACCCGACGAGCGGTTCTCTGTGAGAGCTCTTGGGAG 13706

QY 301 GCGTTGAGCGGCGGGAATCGATCCGCTGTCCGTACGCGGGTCCGCTACGCGGCGTCTTC 360
 DB 13707 GCGCTCGAAGCGGCGGCATCACTCTTCCAGCTGGCGGGCAGTCCGACCGGTGTGTT 13766

QY 361 GCGGGGCGCTCTCTTCCACTACGCGCGCGTATGGACACCGCTCGTCGAGGCGGC 420
 DB 13767 TTCGCGATGCGAACAGGACTACGCGCGCGGCGGCGGACGTCGCGTCCGAG----- 13820

QY 421 CGGACGTCGAGGCGCACATCTCACCGGTACACGCGGAGCGTCTCTGCGGCGGTATC 480
 DB 13821 -----CTGAGGGTACTCTGTCTACCGGCTCAATCTCAGAGTCTGCTTCGGGGGTGT 13874

QY 481 GCCTACAGTTCGCGCTGGAAGCGCGCGCATCACCGTGGACACGCGGTGCTCGGCATCG 540
 DB 13875 GCTTACACGTTCCGTTTCCAGGGGCTCGGTTGACGTTGATACGCGGTTCGTCGTCG 13934

QY 541 CTCGTGACGCTGATCTGGCGTGCAGTCTGCGGTGCGGTGAGTGCACGCTCGCGCTG 600
 DB 13935 TTGGTGGCGTTCATCTGCGGTGCGAGGGGTTGCGCGGGGTGAGTGTTCGTTGCGT 13994

QY 601 GCGCGGCGCTCTCGTCTATGTCACCTCCGATGTTTCATCGAGTTCCTCCGCGGAGCGC 660
 DB 13995 GTGGGTGGGTTGACGCTGATGCTGTCGCGGTGACGTTTACGACGTTTCACTCGGACGCG 14054

QY 661 GGGGTGTCGCTGGAACGCGAGGTGCAAGGCGTACTCGGCTGCGAGCGGACCGGCTGG 720
 DB 14055 GGTGTTGCTGGATGGGCGGTGCAAGGCTTCGCGGCTTCGCGGATGTTTGGTGT 14114

QY 721 GCGAGGCGCTCGGATGCTGTTGTGAGCGGTTGTCCGATCGCGGTGCGGCTGGGCGAT 780
 DB 14115 GCGAGGGGTGCGGTGTGTTGTGAGCGGTTGTCCGATCGCGGCGGTGCGGTAT 14174

QY 781 CGGGGTGTCGCGGTGCTAGCGGCGAGTGGCGTCAACGAGACGCTGCGTCAATGGGCTG 840
 DB 14175 CGGGGTGTCGCGGTGCTGGGGGAGTGGCGTCAATCAGATGGTGGTCAATGTCGTCG 14234

QY 841 ACGGCGCGAAGCGTCCGCTCAGGAGCGGTTGATCCGCGAGCGGTGCGGAGCGGGG 900

DB 14235 GCGGCGCGCAATGGTCCGTTCGACGAGCGGGTATCCGTGCGCGGTGGCTACCGGGT 14294
 QY 901 TTGTCGCTGGCGATGTGATGTGTGTGAGGGGACCGGACCGGACACGACGCTGGGTGAT 960
 DB 14295 CTGGCTCTCGCGATGTGATGTGTGTGAGGGCGATGGACCGGACCGGTTGGGTGAT 14354
 QY 961 CCGATCGAGGACGAGCGGTGCTGCCACGTACGGGCGAG---CGGGCGGTTGACAGGCGG 1017
 DB 14355 CCGATCGAGGCTCAGGCGTTGCTGCGGACGTATGGCGAGGGTCTATACCACTGGGCTCCG 14414
 QY 1018 CTGTGGCTGGGTTCTCTGAAGTCCAAACATCGGGGACACCATGGCTGCCCGGGTGGGT 1077
 DB 14415 GTGTGGCTGGGTTCCGTGAAGTCTGAACATCGGGGATACGACGCGCGCGCGTGGCT 14474
 QY 1078 GGGGTCAATCAAGATGTGATGCGGTTGCGGAGGGGTTGTCGCCGCGGACGTTGCATGTG 1137
 DB 14475 GGTGTGATGAAGATGTTGCTGCGTTGGTTCGGGTTGTGTGCCGAAGACGTTGCATGTG 14534
 QY 1138 GATAAGCGCTCGCCGACGTGACTGTGTCGCGGGGCGGTCGCGTGTGACGAGGCG 1197
 DB 14535 GATGAGCGCTCACCGCATGTGACTGCTGCGCTGTGAGGTGAGTTGCGGCTTGAGGCG 14594
 QY 1198 GTGCGTGGCGGGGACGCGGACAGGCGGTTGCGGCGGGCGGAGTGTCTGCTGTTCCGG 1257
 DB 14595 GTGCGGTGTGTCGCG-----GGTGGCGGGTGTGCGCGGGTGTGTGTGCTGTTCCGG 14648

QY 1258 ATCGGCGCACGAATGCGCATGTGATTGTTGAGGAGGCGCGCGCG 1302
 DB 14649 ATCAGCGGACGAATGCGCATGTGATCGTTGGAGGAGGCGCTGCG 14693

RESULT 12
 AAT80413
 ID AAT80413 standard; DNA; 43280 BP.
 XX
 AC AAT80413;
 XX
 DT 27-FEB-1998 (first entry)
 XX
 DE Ty lactone synthase gene cluster.
 XX
 KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
 KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
 XX
 OS *Streptomyces fradiae*.
 XX
 FH Key Location/Qualifiers
 CDS 816..14243
 FT /*tag= a
 FT /transl_except= (pos: 816..818, aa: Met)
 FT /note= "ORF1 encodes protein shown in AAW22601"
 CDS 14351..19945
 FT /*tag= b
 FT /transl_except= (pos: 14351..14353, aa: Met)
 FT /note= "ORF2 encodes protein shown in AAW22602"
 CDS 20010..31199
 FT /*tag= c
 FT /transl_except= (pos: 20010..20012, aa: Met)
 FT /note= "ORF3 encodes protein shown in AAW22603"
 CDS 31232..36067
 FT /*tag= d
 FT /note= "ORF4 encodes protein shown in AAW22604"
 CDS 36249..41774
 FT /*tag= e
 FT /note= "ORF5 encodes protein shown in AAW22605"
 XX
 EP791655-A2.
 XX
 PD 27-AUG-1997.
 XX
 PF 19-FEB-1997; 97EP-0301056.
 XX


```
QY 721 GCGAGGCGTCCGATGCTGTTGGTGGAGCGGTTCTCGATCGGTGGCGTGGGGCAT 780
    |||||
Db 8978 GCGAGGCGTGGGTGCTGCTGCTGGTGGAGCGGCTTTCGACGCGCGCCGCAACGTCAT 9037
    |||||
QY 781 CGGGTCTGGCGGTGTGACGGCACTGCGGTCAACAGGACGCTGCTGCAATAGGCTG 840
    |||||
Db 9038 CGGGTCTGGCGGTGTGCGGGGACGTGCGGTCAATCAGGACGCTCGGACCAATGGCTG 9097
    |||||
QY 841 ACGGCGCCGACGCTCCGCTCAGGACGCGGTGATCCGACGCGCTTGGCGACGCGGG 900
    |||||
Db 9098 ACGGCGCCGAGTGGTCCGCGCAGCAGCGGTGATCCGTGAGGCGCTTGGCTGATGCGGG 9157
    |||||
QY 901 TTGTCGCTGGCGATGTGGATGTGTGGAGGGGACGCGGACCGGCAACGCTGGGTGAT 960
    |||||
Db 9158 CTGGTCCCGCGCGACGTGGATGTGTGGAGGGCGACGCTACGGGACGCGCTGGGTGAT 9217
    |||||
QY 961 CCGATCAGGACACAGCGGTTCTCGCACGTACGGGACGCGGCGCGGTGACAGGCGCGT 1020
    |||||
Db 9218 CCGATCAGGCGGGTCTGCTGCTGGCCACGTACGCGGCGGGAGCGGGTCGCGCATCGGTTG 9277
    |||||
QY 1021 TGCTCGGGTCTCTGAAGTCCAAATCGGSCACACCATGCTGCGCGGCTGTGGGTGG 1080
    |||||
Db 9278 TGCTCGGGTCTGTTGAAGTCAACATCGGSCATGCGGCGCGCTGCGGTGTGGGTGGT 9337
    |||||
QY 1081 GTCATCAAGATGTGATGCGTGTGCGGAGGGGCTGTTCCGCGGACGTTGTCATGTGGAT 1140
    |||||
Db 9338 GTGATCAAGTGTGTCAGGGGATGCGGCATGGTCTGTCGCGGACGCTGTCATGTGGAT 9397
    |||||
QY 1141 AAGCGTCCGCGCAGTGAGACTGCTGCGCGGGGCGGTGCGGTGTCGACGAGGCGGTG 1200
    |||||
Db 9398 GCGCGCTCGTCAAGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCG 9457
    |||||
QY 1201 CCGTGGCGGGGACCGCGCAGCGGCTTCCGCGCGGCGGAGTGTGCTGTCGGATC 1260
    |||||
Db 9458 TCGTGGCGCGG---CGGGTGGAGCGGGTTCGCGCGGCGCGCGGTGTCGGCGTTCGGGGT 9514
    |||||
QY 1261 GCGCGCACCAATCGCATGTGATTTTGGAGGAGGCGCGCGCG 1302
    |||||
Db 9515 AGCGGACCAACGCCCATGTGTTCTTGGAGGAGCGCGCGCG 9556
    |||||

RESULT 11
AAA09469
ID AAA09469 standard; DNA; 50937 BP.
XX
AC AAA09469;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus oleandomycin gene cluster.
XX
KW oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS Streptococcus antibioticus.
XX
FH Key Location/Qualifiers
FT CDS 152..1426
    /tag= a
    /label= oleI
FT CDS complement (1528..2637)
    /tag= b
    /label= oleN2
FT CDS complement (2658..4967)
    /tag= c
    /label= oleR
FT CDS 5772..18224
    /tag= d
    /label= ORF1
    /product= 8,8a-deoxyoleandolide_synthase_1
```

```
FT CDS 18267..29717
    /tag= e
    /label= ORF2
    /product= 8,8a-deoxyoleandolide_synthase_2
    29787..40346
    /tag= f
    /label= ORF3
    /product= 8,8a-deoxyoleandolide_synthase_3
    40625..41830
    /tag= g
    /label= oleP1
    41878..43158
    /tag= h
    /label= oleG1
    43163..44443
    /tag= i
    /label= oleG2
    44433..45173
    /tag= j
    /label= oleM1
    45251..46411
    /tag= k
    /label= oleY
    46491..47714
    /tag= l
    /label= oleP
    complement (47808..49517)
    /tag= m
    /label= oleB

WO200026349-A2.
11-MAY-2000.
XX
XX 22-OCT-1999; 95WO-US244478.
XX
XX 29-OCT-1998; 98US-0106100.
XX 16-FEB-1999; 99US-0120254.
XX (KOSA-) KOSAN BIOSCIENCES INC.
XX
XX Betlach MC, Shah SK, McDaniel R, Tang L;
PI WPI; 2000-365602/31.
DR P-PSDB; AAY92707, AAY92708, AAY92709.
XX
XX Recombinant DNA compound encoding oleandolide polyketide synthase for
XX synthesizing polyketides comprising a coding sequence for a domain of a
XX loading module or any one of extender modules
XX
XX Disclosure; Page 14-26; 86pp; English.
XX
XX This is part of the Streptococcus antibioticus oleandomycin gene cluster.
XX The oleandolide polyketide synthase (PKS), also known as
XX 8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
XX (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
XX "modular" enzyme, where each ORF encodes 2 extender modules and
XX the first ORF also encodes the loading module. Each module is composed
XX of at least a ketosynthase (KS), acyl-transferase (AT) and an
XX acyl carrier protein (ACP) domain. The oleandolide PKS loading module
XX contains an inactivated KS, called KS-Q, where Q is the abbreviation for
XX glutamine, present instead of the active site cysteine required for
XX activity. The large multifunctional PKS enzymes catalyze the biosynthesis
XX of polyketide macrolactones through multistep pathways involving cycles
XX of decarboxylative condensations between acylthioesters followed by cycles
XX of varying beta-carbon processing activities. The macrolide product of
XX the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
XX glycosylation to yield oleandomycin, an antibacterial polyketide. The
XX invention concerns an isolated recombinant DNA compound, comprising a
XX coding sequence for a domain of loading module or any one of extender
XX modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
XX promoter. Also discussed are recombinant oleandolide PKS in which the
XX module 1 KS domain is inactivated by deletion or other mutation. In
```

Db 9278 TGGCTCGGTGTTGAAGTCGAACATCGGGCATGCGACGGCGCTCCGGGTGTGGGTGT 9337
 Qy 1081 GTCATCAAGATGATGATGCGTTGCGGAGGGGGTGTTCGCGGACGAGTTCATGCGAT 1140
 Db 9338 GTCATCAAGTGTGTCAGGGATGCGCATGGTCTGTCGCGGACGCTGATGCGAT 9397
 Qy 1141 AAGCCGTCCGCGAGTGATGCTGTCGCGGGGGCGGTGCGGTGCTGACGAGCGGTG 1200
 Db 9398 GCGCCGTCTCAAGTGAAGTGGCTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTC 9457
 Qy 1201 CCGTGGCCGGGACCGCGACAGGCGGTTCGCGGCGCGGAGTTCGTCGTCGCGATC 1260
 Db 9458 TCGTGGCCCGCG--CGGTGAGAGCGGTGCGCGCGCGCGCGCGTTCGCGTTCGCGGTG 9514
 Qy 1261 GCGCGGACCAAGTCGATGATTTGGAGGAGCGCGCGCG 1302
 Db 9515 AGCGGGACCAAGCCCATGTGCTCTGGAGGAGCGCGCG 9556

RESULT 10

AAT80414

ID AAT80414 standard; DNA; 44377 BP.

XX AC AAT80414;

XX DT 27-FEB-1998 (first entry)

XX DE Platenolide synthase gene cluster.

XX KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;
 KW platenolide synthase gene cluster; platenolide production; strmg gene;
 XX polyketide; tyllactone synthesis; antibiotic; tyllactone; ss.
 XX OS Streptomyces ambofaciens.

XX FH Key Location/Qualifiers
 FT CDS 350..14002
 FT /tag= a
 FT /transl_except= (pos:350..352, aa:Met)
 FT /note= "ORF1 encodes protein shown in AAW22606"
 FT CDS 1406..20036
 FT /tag= b
 FT /note= "ORF2 encodes protein shown in AAW22607"
 FT CDS 20110..31284
 FT /tag= c
 FT /transl_except= (pos:20111..20113, aa:Met)
 FT /note= "ORF3 encodes protein shown in AAW22608"
 FT CDS 31329..36071
 FT /tag= d
 FT /note= "ORF4 encodes protein shown in AAW22609"
 FT CDS 36155..41830
 FT /tag= e
 FT /note= "ORF5 encodes protein shown in AAW22610"

XX EP791655-A2.

XX 27-AUG-1997.

XX PF 19-FEB-1997; 97EP-0301056.

XX PR 22-FEB-1996; 96US-0012078.

XX PA (ELIL) LILLY & CO ELI.

XX PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;

XX DR WPI; 1997-418046/39.

XX DR P-PSDB; AAW22606-W22610.

XX PT DNA encoding Streptomyces fradiae tyllactone synthase domain - for
 PT production of tyllactone-related polyketide compounds
 XX Example 2; Pages 110-134; 220pp; English.

XX This sequence represents the platenolide synthase gene cluster of the
 CC invention. This sequence is referred to as the strmg gene, and was
 CC isolated from Streptomyces ambofaciens. This sequence encodes the
 CC multi-functional proteins which direct the synthesis of the polyketide
 CC platenolide. Platenolide is the basic building block of the macrolide
 CC antibiotic spiramycin. This sequence was used along with the tylG gene
 CC (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
 CC gene is the tyllactone synthase gene cluster of the invention. The tylG
 CC sequence was isolated from Streptomyces fradiae, and encodes
 CC multifunctional proteins which direct the synthesis of the polyketide
 CC tyllactone. Tyllactone is the basic building block of the antibiotic
 CC tyllactin. The hybrid sequence can be used to transform S. ambofaciens
 CC lacking the strmg ORF1 sequence, or S. fradiae lacking the tylG ORF1
 CC sequence, so that they can produce polyketides. The DNA sequence can be
 CC modified so as to alter the type of carboxylic acids incorporated, the
 CC number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tyllactone-related
 CC polyketides.
 XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
 Query Match 54.1%; Score 704.8; DB 18; Length 44377;
 Best Local Similarity 72.6%; Pred. No. 3.3e-112;
 Matches 945; Conservative 0; Mismatches 342; Indels 15; Gaps 2;
 Qy 1 GACCGCATCGCCATCATCGGCATGCGTTCGCGGCGGAGTCCGGTCCCGAAG 60
 Db 8270 GAGCCGATCGCGATCGTGGCATGCGTTCGCGGCGGAGTCCGGTCCCGAAG 8329
 Qy 61 GACCTGTGGAGCTGGCGCCCTCGCGGCGGAGACGCCATCGGGCCGTTCCCAACCGACCG 120
 Db 8330 GACCTGTGGAGCTGGCGCCCTCGCGGCGGAGACGCCATCGGGCCGTTCCCAACCGACCG 8389
 Qy 121 GGATGGCCCAAGAAACAGCGTTCACGCCAGAGACCCAGACGCCCGGCGGAGTCTATCCG 180
 Db 8390 GCGTGGGACGTCGAGGGGCTGTACAGCCGCGGAGCGGGGTCGCGGCGGAGGACTATGTA 8449
 Qy 181 CAGGAGGCGGGTTCCTTCAGAGCGCGGCACTTCGACGCCGGTCTTCGCGGATCACT 240
 Db 8450 CCGGAGGCGGGTTCCTTCAGAGCGCGGCACTTCGACGCCGGTCTTCGCGGATCACT 8509
 Qy 241 CCAGTGTAGGCACTGGCGATGATCCGCGAGCGCGGCTGCTCTGGAGACGTCCTGGGAG 300
 Db 8510 CCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8569
 Qy 301 GCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 Db 8570 GCGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8629
 Qy 361 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 Db 8630 AGCGGCACTCTCCAGCAGGACTAC-----GGACCCAGCTGGGGGACGCGCGC 8677
 Qy 421 CCGGACGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 Db 8678 GACACCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8737
 Qy 481 GCCTACAGCTTCGGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 Db 8738 GCCTATGCGTTGGGCTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8797
 Qy 541 CTCGTGACGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 Db 8798 TTGGTGGCGTTCGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8857
 Qy 601 GCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 Db 8858 GCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8917
 Qy 661 GCGGTGTGCGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 Db 8918 GCGGTGTGCGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8977

XX Polyketide synthase; oxidative modification; metabolite; antibiotic; antitumor; pimaricin, 88.
KW Streptomyces natalensis.
XX Key Location/Qualifiers
XX CDS 1..20394
XX /tag= a
XX /product= "polyketide synthase"
XX W0200077222-A1.
XX 21-DEC-2000.
XX 14-JUN-2000; 2000WQ-EP06227.
XX 14-JUN-1999; 99EP-0201893.
XX (STAM) DSM NV.
XX Martin JF, Aparicio JF, Colina AJ;
XX WPI; 2001-080693/09.
XX P-PSDB; AAB31558.
XX New polynucleotides encoding enzymes involved in the biosynthesis of pimaricin, useful for modifying the biosynthesis of pimaricin and in the synthesis of new compounds -
XX Disclosure; Page 53-80; 116pp; English.
XX The present sequence encodes a polyketide synthase which is associated with the biosynthesis of pimaricin. The polyketide synthase polypeptide is useful for the oxidative modification of a methyl group of a suitable compound, e.g. a bioactive compound including a secondary metabolite, antibiotics and anticancer agents. Recombinant cells comprising the gene are useful for the production of pimaricin. The polyketide synthase polynucleotide may be over expressed in Streptomyces, leading to an increase in the biosynthesis of pimaricin, as a source of primers for amplification reaction and as probes.
XX
SQ Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;
Query Match 54.5%; Score 710; DB 22; Length 20394;
Best Local Similarity 73.7%; Pred. No. 4.5e-113;
Matches 956; Conservative 0; Mismatches 315; Indels 27; Gaps 3;
QY 1 GACCCGATCCCATCATCGGCATGCGATGCGCTTTCCCGGGCGGAGTCCGGTCCCGGAAG 60
DB 9895 GACCCGATCGTATCGTGGCATGAGCTGCGCTACCCGGGCGGCTCGGCTCGCCGAG 9954
QY 61 GACCTGTGGAGCTGGCGGCTCGGGCGGAGAGCCATCGGGCGGTTCCCGACCGACCG 120
DB 9955 GACCTGTGGCGCTGTGTGCGAGGGCGTGGACGCGGTGTCGACTTCCCGACCGACCGT 10014
QY 121 GGATGGCCACGGAACAGCTACGCCAGGACCCACGACGCCGCGACGCTCTATCCG 180
DB 10015 GGATGGAGCTGGAGAGCCCTTACAGCCCGGAGCGCTCGGACCTCGTACACC 10074
QY 181 CAGGGAGGCGGGTTCTTACAGCGCGGCGCACTTCGACCGCGGCTTCTTCGGAATCAGT 240
DB 10075 CGCTCCGTTGATTCCTCCAGAGGCGCGGAGTTCGACCCGATTTCTCGGGATGAGC 10134
QY 241 CCAGGTGAGCACTGGCGATGATTCGCGACGAGCGGCTGCTGTGAGAGCTCTGGGAG 300
DB 10135 CCGCGGAGCGCTGGCGACCGACCGCCAGCAGCGGCTGCTGTGGAGAGCACTGGGAG 10194
QY 301 GCGTTCCAGCGGGGGAATCGATTCGCTGCGGTACCGGGTCCCGTACGGGCGCTTC 360
DB 10195 GCATCCAGCGCAGGGGATCGACCCCGGCTGCTGCGGGGCGAGCCGCTACGGGCGCTTC 10254
QY 361 GCGGGCGCCCTCTCTTCGACTACGGCCCGCGCTATGACACCGCGCTGCTCGAGGGCGCC 420

DB 10255 GCGGGGTGATGTATACCGACTACGGGACCTCTCTCGTGGCGACAGT----- 10303
QY 421 GCGGACGTGGAGGCCACATCTTACCGGTACACGGGACGGGCGGTCTTCTGGGCGGTATC 480
DB 10304 -----TCGAGGGCTACCGCAGCAACCGGAGCGCGCAGCATCGCTCCGGCGGGTC 10356
QY 481 GCCTACAGCTTCGGGCTGGAAGGGCGGCGCATACCGTGGACACGGGGTGTCTCGGCATCG 540
DB 10357 TCGTACACTTCGGTTTCGAGGGTTCGGGCGTACGGTGGACCGGCATGCTCTGCTGCC 10416
QY 541 CTCGTGACGCTGTCATCTGGCGTGCAGTCTGCGGTTCGGGTGAGTGCACGCTCGCGCTG 600
DB 10417 CTGGTGGCCCTGCACTGGGCGGCGAGTCTGCTGGCTCGGCGAGTGTCTGCTCGCGTC 10476
QY 601 GCCGGGCGGCTCTCGGTCTATCTCCACCTCGGCATGTTTCATCGAGTTCTCCGGCAGCGC 660
DB 10477 GCGGGCGGTGTGACGGTGTATGTCCACCGACGAGTTTCTGTCGAGTTCTTCGGGCAACGC 10536
QY 661 GGGCTGTCTGGTGGACGCGCAGGTGCAAGGCGTACTCGGCTGCAGCGCAGCGCACCGGCTGG 720
DB 10537 GGAATGTCTGGCGGACGGCGCTGCAAGGCGTTCGCCGATGCGGCGACGGGCTCGGCTGG 10596
QY 721 GCGGAGGCGCTCGGATGCTGTGGTGGAGCGGTGTCTCGGATGCGGTTCGGCTGGGGCAT 780
DB 10597 GCGGAGGCGCTCGGATGCTGTGGTGGAGCGGTGTCTCGGACGCGCGCGCAACCGGCGAC 10656
QY 781 CCGGTGTCTGGCGGTGTGTACGGGCGAGTCTCGGTCAACAGGACGGTGTCTGCAATGGGGTG 840
DB 10657 CCGGTGTCTGGCGGTGTGTGCGGCGAGTCTGCGGTGAACAGGACGGTGTCTCAATGGTCTG 10716
QY 841 ACGGCGCGCAACGGTCTCGGCTCAGGAGCGGTGTATCTCGGCGAGGCTTTCGCGCAACCGCGGG 900
DB 10717 ACGGCGCGCAACGGCGCGCGCGCGCGCGGTGTATCTCGGCGAGGCTTTCGCGAGTCCGGGG 10776
QY 901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCGACGGGACCGGACGCTGGGGTGTAT 960
DB 10777 CTGTCCGCGCGGATGTGGACGCGGTGGAGGCGCACGCTACGGGTACGACGCTGGGCGAT 10836
QY 961 CCGATCGAGGACACAGGCGGTCTGCTCCACGCTACCG--GCAGCGGGCGGTGACAGGCGG 1017
DB 10837 CCGATCGAGGCGCGCGCGCTGCTGCCACGATATGGCCAGGAGCGACCTGAGAGCGCGCG 10896
QY 1018 CTGTGGCTGGGGTCTCTCAAGTCCAAATCCGGGACACCATGGCTGCGCGGGTGTGGGT 1077
DB 10897 TTGCTGTGGGTCTGGTCAATCCAAATCCGATCGTATGCGCAGGCGGCTTTCGGGTGTGGCG 10956
QY 1078 GGGGTCTCAAGATGGTGTATGGCTGGGAGGGGGTGTGCGCGGACGCTTGCATGTG 1137
DB 10957 GGTGTCTCAAGATGGTGTATGGCTGGGAGGGGCGTGTGCTTCCGACGCTGCTATGTG 11016
QY 1138 GATAGCGCTGCGCGCAGGTGGACTGTGTCGGGGGGCGGTGCGGCTCTGACGGAGCG 1197
DB 11017 GATGAACGCTGCTGCTCATGTGCTGAGTGGAGTGGCGGTGCGGCTGCTGACCTCCGAG 11076
QY 1198 GTGCGCTGGCGGGGAGCGCGGAGGCGGTGTCGGCGGGGGGAGTGTCTGCTGTTCGGG 1257
DB 11077 GCGGAGTGGCGCAGGGCGAGS-----GGCGCGCGCGGGGCTCTCTCTCTTCGGC 11130
QY 1258 ATCGCGCGCAGCAATTCGCATGTGATTTTGGAGGAGGC 1295
DB 11131 GTCAGCGGACGCAACCGCATGTGATCTCTGGAGCAGCC 11168

RESULT 8

AAT80415

ID AAT80415 standard; DNA; 13987 BP.

XX

AC AAT80415;

XX

DT 02-MAR-1998 (first entry)

XX

DE Hybrid armG/tylG ORF1.

OS Streptomyces avermitilis.

XX Key Location/Qualifiers
 FH 1..14646
 CDS /tag= a
 FT /product= "AAG65266"
 FT 14824..31422
 CDS /tag= b
 FT /product= "AAG65267"
 XX

WO200162939-A1.

XX

XX 30-AUG-2001.

XX

XX 23-FEB-2001; 2001WO-JP01381.

XX

XX 24-FEB-2000; 2000JP-0047405.

XX

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX

XX (KITA) KITASATO INST.

XX

XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX

XX WPI; 2001-582053/65.

XX

XX P-PSDB; AAG65266, AAG65267.

XX

XX New modified avermectin aglycone synthase derived from Streptomyces

XX

XX avermectilis used in production of 22,23-dihydroavermectin B1a used in

XX

XX drugs and pesticides -

XX

XX Disclosure; Page 103-149; 257pp; Japanese.

XX

XX The present invention relates to the production of modified derivatives

XX

XX of avermectin aglycone synthase (AMS) derived from Streptomyces

XX

XX avermitilis. The activity of an acyl carrier protein (ACP),

XX

XX beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),

XX

XX beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl

XX

XX reductase (ER) and/or thioesterase (TE) domain may be reduced or

XX

XX suppressed. The process can be used in the production of drugs, veterinary

XX

XX drugs and pesticides. The present sequence is a fragment of the S.

XX

XX avermitilis genome.

XX

XX

XX

XX

XX

XX

XX

XX

SQ

Query Match 72.6%; Score 944.8; DB 22; Length 31422;
 Best Local Similarity 83.6%; Pred. No. 3.1e-153;
 Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;
 1 GACCCGATCGCCATCATCGGCATGCGATCCGTTTCCCGGCGGAGTCCCGTCCCGAAG 60
 14935 GAGCCGATTGCGATTGCGGGATGGCGTGTCTGTTTCCGGGTGGGTGTTTCGCGGAG 14994
 61 GACCTGTGGAGTGGCCCTCGGGCGAGAGCGCATCGGCGCGTTCCCGACCGACCGC 120
 14995 GAGTGTGTGGAGCTGTGTCGTGGGTGGGATGCGATTGGTGAATTTCCGCGCGTCCG 15054
 121 GGATGCCCCACGAAACAGCGTCAAGCCAGGACCCCGACCGCGGACGTTCTATCCG 180
 15055 GGTGGGATCTGAGGGGTGTTGTTGATTCGACCTTGACCGGTGCGGGACGTCGTACGG 15114
 181 CAGGGAGGCGGGTTCCTTCACAGCGCGGCATCTTCGACGCGCGGTCTTTCGGAATCAGT 240
 15115 CGGTATGCGGGTGTGTTGATGAGGCGGGGAGTTCGATGCGGACTTCTTCGGGATCAGT 15174
 241 CCACGTGAGCACTGGCGATGATTCGAGCGAGCGGCTCTGCTGAGACGTCCTTGGAG 300
 15175 CCGCGTGAAGCGTGGCGATGATTCGAGCGAGCGGTTTGTGTCGAGACGTCGTGGGAG 15234
 301 GCGTTCGAGCGGCGGGAATCGATCCGTCGTGTCGAGCGGCTCCGTCACGGCGCTTC 360
 15235 GCGTTCGAGCGGCGGGAATCGATCCGTCGTGTCGAGCGGCTCCGTCACGGGCTTC 15294
 361 GCGGGCGCCCTCTCTTCGATTCAGCTACGCGCCCGCGTATGGACACCGCGTCGTTCGAGGGCGCC 420

Db 15295 GCCGGGTGATGTACACGACTACGGATCCCGCCTGGGTACCAT-----CCCC 15342
 QY 421 GGGACGTGGAGGGCCACATCTCTACCGGTATACACGGGAGCGTCTCTGTGCGGCCGTATC 480
 Db 15343 GAGGATTCGAGGGCTATATCGGCAACGATAGCGGGGCGCGCTCGCTGCGGCGCGCTC 15402
 QY 481 GCCTACAGCTTCGGCTCGGAAGGGCCGGCGATCACCTGTGGACACAGGGGTCTCGGCATCG 540
 Db 15403 GCCTACAGCTTCGGCTCGAGGGCCCTGCCGTCTCGGTGGACACAGCATGTTCTCGTCTG 15462
 QY 541 CTCGTGACGCTGCTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 600
 Db 15463 TTGGTGGCGTGCATCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 15522
 QY 601 GCGGCGGCGCTCTCGGTGCTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 660
 Db 15523 GCGGCGGCGTGTGACGCTGATGTCGACCCCGCCGCTCTTCTGTCGATGTTCTTTCACGCCAGCC 15582
 QY 661 GGGCTGTGCTGCGGTGCGGAGTGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 720
 Db 15583 GGACTGTGCTGCGGTGCGGAGTGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15642
 QY 721 GCGGAGGCGCTGCGGTGCTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 780
 Db 15643 GCGGAGGCGCTGCGGTGCTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15702
 QY 781 GCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 840
 Db 15703 GCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15762
 QY 841 AGCGGCGGCGGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 900
 Db 15763 AGCGGCGGCGGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15822
 QY 901 TTGTCGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 960
 Db 15823 TTGTCGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15882
 QY 961 CGGATCGAGGCGGCTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 1020
 Db 15883 CGGATCGAGGCGGCTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 15942
 QY 1021 TGGCTGGGCTCTCTGAAGTCCAAACATCGGCGCACACCATGCGGTGCGGTGCTGCGGTGCTGCGGTG 1080
 Db 15943 TGGCTGGGCTCTCTGAAGTCCAAACATCGGCGCACACCATGCGGTGCGGTGCTGCGGTGCTGCGGTG 16002
 QY 1081 GTCTCAAGATGCTGATCGGCTTTCGCGGAGGGGGTTCGCGCGGACGCTTTCGATGCTGAT 1140
 Db 16003 GTCTCAAGATGCTGATCGGCTTTCGCGGAGGGGGTTCGCGCGGACGCTTTCGATGCTGAT 16062
 QY 1141 AAGCGCTCGCGGAGTGGATGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 1200
 Db 16063 GAGCGCTCGCGGAGTGGATGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 16122
 QY 1201 CCGTGGCGGCGGAGCGGCGGAGGCGGTTCGCGGCGGCGGCGGTTCGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 1260
 Db 16123 CCGTGGCGGCGGAGCGGCGGAGGCGGTTCGCGGCGGCGGCGGTTCGCGGTGCTGCGGTGCTGCGGTGCTGCGGTG 16182
 QY 1261 GCGCGCACGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 1302
 Db 16183 AGTGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 16224

RESULT 7

AAFP24892

ID AAF24892 standard; DNA; 20394 BP.

XX

XX AAF24892;

XX

XX 20-APR-2001 (first entry)

XX

XX Pimaricin biosynthesis associated polyketide synthase gene.

/note= "avermectin aglycon synthase protein"

CDS 14824..31422

/tag= b

/note= "avermectin aglycon synthase protein"

WO200050605-A1.

31-AUG-2000.

23-FEB-2000; 2000WO-JP01041.

24-FEB-1999; 99JP-0046961.

(KITA) KITASATO INST.

Omura S, Ikeda H;

WPI; 2000-565458/52.

P-FSDB; AAB23751, AAB23752.

Avermectin aglycone synthase DNA and proteins encoded by all or part of it for the production of avermectin and its derivatives for drug and agrochemical use

Claim 2; Page 134-203; 314pp; Japanese.

The present sequence represents DNA which encodes avermectin aglycon synthase proteins. Also described are: (1) polypeptides encoded by all or part of the DNA; (2) expression vectors containing the DNA; (3) host cells transformed by the vectors; (4) preparation of the polypeptides by culture of the transformants; (5) preparation of avermectin aglycon or its derivatives by culture of transformed avermectin-producing microorganisms; and (6) oligonucleotides of 5-60 bases in length containing sense or antisense sequences from the avermectin aglycon synthase DNA. The enzymes are useful for the production of modified forms of avermectin and of the intermediates in its biosynthesis, for use as drugs, veterinary drugs and agrochemicals.

Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

Query Match 72.6%; Score 944.8; DB 21; Length 31422;
Best Local Similarity 83.8%; Pred. No. 3.1e-153;
Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;
1 GACCCGATGCCATCATCGCATGCGATGCGGTTCCCGGGCGAGTCCGGTCCCGAAG 60
14935 GAGCCGATTCGATTTGGGGATGGCGTGTCTTTCCGGTGGGTGTTCGCCGGAG 14994
61 GACCTGTGGAGTGGCCGCTCGGCGGAGAGCCATCGGGCGTTCCCAACCGACCGC 120
14995 GAGTTGTGGAGCTGGTTCGCTCGGTGGGATGCGATTGGTGAATTCGGCGCGTCGG 15054
121 GGATGCCCAACGACAGCGTACGCCAGGCCAGGCCACCGAGCCCGGACGTTCTATCCG 180
15055 GGGTGGGATCTGGAGGGGTGTTTGTATGAGGGGGGGAGTTCGATCGCGACGTCGTAC 15114
181 CAGGGAGGGGGGTTCCTTACAGACGGCGGCACTTCGACGCGCGTTCCTCGGAATCAGT 240
15115 CGGTATGGCGGGTTCCTTATGAGGGGGGGAGTTCGATCGCGACTTCCTCGGGATCAGT 15174
241 CCACGTGAGGCACTGCGGATGGATCCGACAGCGGCTGCTGTGGAGAGCTCTCGGGAG 300
15175 CCGCGTGAGCGGTGGCGATGGATCCGACAGCGGTTGTTGTGGAGAGCTGTGGGAG 15234
301 GCGTTCGAGCGGGGGGATCGATCCGCTGCTGGTACGCGGGTCCCGTACGGCGCTTC 360
15235 GCGTTCGAGCGGGGGGATCGATCCGCTGCTGATCGGTGGCTCCCGTACGGGTGCTTC 15294
361 GCGGGCGCCCTCTCTTCGACTACGCGCGCGTATGGACACCGCGTCTCGGAGGCGCC 420
15295 GCGGGGGTGTATACACGACTACGATCCCGCTGGGTACCAT-----CCCC 15342
421 GCGGACGTGGAGGGGCCACATCTCTACCGGTACACCGGCGAGCGTCTCTCGGGCGTATC 480

Db 15343 GAGGATTCGAGGGGCTACATCGGCAACGGTAGCGCGCGCGCTCGCTCGCGCGCGCTC 15402
Qy 481 GCCTACAGCTTCGGGCTGGAGAGGGCGGCGATACCGTGGACACGGGGTGTCTCGGCATCG 540
Db 15403 GCCTACAGCTTCGGGCTCGAGGGGCTCCCGTCTCGGTGGACACGGCATGTTCTGCTG 15462
Qy 541 CTCGTGACGCTGCTATCTGGCGTGGCAGTCTGCGGTGGGTGAGTGACGCTCGCGCTG 600
Db 15463 TTGGTGGCGCTGCTATCTGGCGTGGCAGTCTGCGGTGGGTGAGTGACGCTCGCGCTG 15522
Qy 601 GCGGGCGCGCTCTCGGTGCTATGTCACCTCGGCGATGTTTCATCGAGTTCCTCCGGAGCGC 660
Db 15523 GCGGGCGGTGACAGGTGATGTCGACCCCGCACCTCTTCGTGAGTTCCTACGCCAGCGC 15582
Qy 661 GGGCTGTCTGGTGGAGCGGAGGTGCAAGGCTGCTCGGTGACGCGCGGACCGCACCGGCTGG 720
Db 15583 GAGCTGTCTGGTGGAGCGGCGCTGCAAGTCTTCCTGGCGGTGGAGCGGACCGCACCGGCA 15642
Qy 721 GCGGAGGCGCTCGGATGCTGTTGGTGGAGCGGTTGTGCGATGCGGTGCGGTGCGGCGAT 780
Db 15643 GCGGAGGCGCTCGGATGCTGTTGGTGGAGCGGTTGTGCGATGCGGTGCGGTGCGGCGAT 15702
Qy 781 CGGCTGTCTGGCGGTGCTGACGCGCGCATGTCGGTCAACGAGGAGCGGTGCGTCAATGCGGTG 840
Db 15703 CGGCTGTCTGGCGGTGCTGCGCGCGCATGTCGGTCAATCAGGAGCGGTGCGTCAATGCGGTG 15762
Qy 841 ACGGCGCGAAGCGTCCGCTCAGGAGCGGCTGATCCGCGAGGCGTTCGCGAACCGCGGG 900
Db 15763 ACGGCGCGAAGTGTGCGCTCAGGAGCGGCTGATCCGCGAGGCGGTTGCGCGAACCGCGGG 15822
Qy 901 TTGTCGTGGCGGATGTCGATGTCGAGGGGCGACGGGACGGGACGACGCTGGGTGAT 960
Db 15823 TTGTCGTGGCGGATGTCGATGTCGAGGGGCGATGCGGACGGGACGACGCTGGGTGAT 15882
Qy 961 CGATCGAGGACAGGCGCTTGTCCGACGATCGGCGAGCGGCGCGGTGACAGGCGCGCTG 1020
Db 15883 CGATCGAGGCGCGAGCGCTTGTCTCCGACGATCGGCGAGCGGCGCGGTGACAGGCGCGCTG 15942
Qy 1021 TGGCTGGGCTCTCGAAGTCCAAATCGGCGACACCATGGCTGCGCGGGTGTGGGTGG 1080
Db 15943 TGGCTGGGATCGGTGAAGTCGAAATCGGCCATGCGCAGGCTGCCCGGGTGTGGGTGG 16002
Qy 1081 GTCTCAAGATGCTGATGCGCTTGGGAGGGGGTGTTCGCGCGGACGCTTGCATGCGAT 1140
Db 16003 GTCTCAAGATGCTGATGCGCTTGGGAGGGGGTGTTCGCGCGGACGCTTGCATGCGAT 16062
Qy 1141 AAGCGCTCGCGCAGGTGGAATGCTTCCGCGGGGGCGGTGCGGCTGCTGACGAGGCGGTG 1200
Db 16063 GAGCGCTCGCGCAGGTGGAATGCTTCCGCGGGGGCGGTGCGGCTGCTGACGAGGCGGTG 16122
Qy 1201 CGGTGGCGGGGACCGGCGAGGGCGGTTGCGGGGGGGAGTGTGCTGCTTCGGGATC 1260
Db 16123 CGGTGGCGGGGACCGGCGAGGGCGGTTGCGGGGGGGAGTGTGCTGCTTCGGGATC 16182
Qy 1261 GCGCGCACGAAATGCGCATGTGATTTTGAGGAGGCGCGCGCG 1302
Db 16183 AGTGGCACGAAATGCGCATGTGATTTTGAGGAGGCGCGCGCG 16224

RESULT 6
AAH79278

ID AAH79278 standard; DNA; 31422 BP.

XX AC AAH79278;

XX DT 04-DEC-2001 (first entry)

XX DE Streptomyces avermitilis coding sequences SEQ ID NO: 2.

XX KW Avermectin aglycone synthase; AAS; avermectin derivative;

XX XX drug production; veterinary drug; pesticide; ds.

XX 06-JUL-1998; 98GB-0014622.
XX (BIOT-) BIOTICA TECHNOLOGY LTD.
XX (PIFZ) PFIZER INC.
XX Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
XX McArthur HAI;
XX WPI; 2000-182117/16.
XX Mutated Type I polyketide synthase containing a polylinker site in an
XX extension module for replacement of a reductive loop sequence, for
XX producing polyketides, e.g. B1 avermectin -
XX Disclosure; Fig 7a-f; 75pp; English.
XX The present sequence is that of DNA encoding the first 2 modules
XX of the avermectin polyketide synthase (PKS) of Streptomyces
XX avermitilis. The invention relates to nucleic acids encoding a
XX Type I PKS such as avermectin in which a polylinker with multiple
XX restriction sites replaces or 1 more PKS genes encoding enzymes
XX associated with reduction. Novel PKS are provided in which in
XX which the reductive loop in a selected module of the type I PKS is
XX replaced with the equivalent segment from the same or different
XX PKS gene cluster or by a mutated or synthetic segment. Vectors and
XX host cells, and methods for producing novel polyketides by
XX culturing host cells are claimed. The polyketides obtained are
XX useful as antibiotics and insecticides. Fermentation products
XX containing C22-C23 dihydroavermectin, ivermectin and B1
XX avermectins are claimed.
XX Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 99.8%; Score 1298.8; DB 21; Length 12381;
Best Local Similarity 99.8%; Pred. No. 9.3e-214;
Matches 1300; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GACCGATCGGCATCGCATCGCATCGCATCGGCGGCGGAGTCCGCTCCCGAAG 60
DB 2258 GACCGATCGGCATCGCATCGCATCGCATCGGCGGCGGAGTCCGCTCCCGAAG 2317
QY 61 GACCTGTGGAGCTGCGCCCTCGGCGGAGAGCGCATCGGCGGCTCCCGACCGACGCG 120
DB 2318 GACCTGTGGAGCTGCGCCCTCGGCGGAGAGCGCATCGGCGGCTCCCGACCGACGCG 2377
QY 121 GATGCGCCACGACAGCGTCACGCCAGGAGCCCGACGAGCGCGGAGCTTCTATCGG 180
DB 2378 GATGCGCCACGACAGCGTCACGCCAGGAGCCCGACGAGCGCGGAGCTTCTATCGG 2437
QY 181 CAGGAGGCGGCTTCTTCCAGACGCGCGCATCTCGACGCGGCTTCTTCCGGAATCAGT 240
DB 2438 CAGGAGGCGGCTTCTTCCAGACGCGCGCATCTCGACGCGGCTTCTTCCGGAATCAGT 2497
QY 241 CCACGTGAGGCACTGCGCATGATCGCAGACGCGCTGCTGCGAGACGCTCTGGGAG 300
DB 2498 CCACGTGAGGCACTGCGCATGATCGCAGACGCGCTGCTGCGAGACGCTCTGGGAG 2557
QY 301 GCGTTCGACGCGGCGGGAATCGATCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTTTC 360
DB 2558 GCGTTCGACGCGGCGGGAATCGATCGCTGTGCTGCTGCTGCTGCTGCTGCTTTC 2617
QY 361 GCGGCGGCGCTCTCTTCTGACTACGCGCGCGCTATGGACACGCGCTGCTGCGAGGCGCGC 420
DB 2618 GCGGCGGCGCTCTCTTCTGACTACGCGCGCGCTATGGACACGCGCTGCTGCGAGGCGCGC 2677
QY 421 GCGGAGCTGAGGCGGCACATCTCTACCGGTACACGCGGAGCGGTCTCTGCTGCGGCGGTATC 480
DB 2678 GCGGAGCTGAGGCGGCACATCTCTACCGGTACACGCGGAGCGGTCTCTGCTGCGGCGGTATC 2737
QY 481 GCCTACAGCTTCGGGCTGGAAGGCGCGGCGATCACGCTGGACACGCGGCTGCTGCGGATCG 540
DB 2738 GCCTACAGCTTCGGGCTGGAAGGCGCGGCGATCACGCTGGACACGCGGCTGCTGCGGATCG 2797

QY 541 CTCGTGACGCTGTCATCTGCGCTGCGAGTCCAGTCCGCTGCGGTGAGTGACGCTGCGCTG 600
DB 2798 CTCGTGACGCTGTCATCTGCGCTGCGAGTCCAGTCCGCTGCGGTGAGTGACGCTGCGCTG 2857
QY 601 GCGGCGGCGCTCTCGGTATGCTTCCACCTCCGTCATGCTGCGGTGAGTGCTTCCGCGGACGCG 660
DB 2858 GCGGCGGCGCTCTCGGTATGCTTCCACCTCCGTCATGCTGCGGTGAGTGCTTCCGCGGACGCG 2917
QY 661 GCGGCTGCTGCGTGGACGCGAGGTGCAAGCGTACTCGGCTGCGAGCGGACGCGGCTGCG 720
DB 2918 GCGGCTGCTGCGTGGACGCGAGGTGCAAGCGTACTCGGCTGCGAGCGGACGCGGCTGCG 2977
QY 721 GCGGAGGCGCTGCGGATGCTGTTGCTGAGACGCGTGTGCTGCGATGCGGTGCGGTGCGGAT 780
DB 2978 GCGGAGGCGCTGCGGATGCTGTTGCTGAGACGCGTGTGCTGCGATGCGGTGCGGTGCGGAT 3037
QY 781 GCGGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 3038 GCGGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3097
QY 841 ACGGCGCGGACGCGTCCGCTCAGGAGCGGCTGATCCGCGAGCGGTGCTGCGGACGCGGCG 900
DB 3098 ACGGCGCGGACGCGTCCGCTCAGGAGCGGCTGATCCGCGAGCGGTGCTGCGGACGCGGCG 3157
QY 901 TTGTCGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 3158 TTGTCGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3217
QY 961 CCGATCGAGGACAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 3218 CCGATCGAGGACAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3277
QY 1021 TGGCTGGGCTCTGAGTCCAAATCGGCGCACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 3278 TGGCTGGGCTCTGAGTCCAAATCGGCGCACACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3337
QY 1081 GTCATCAAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
DB 3338 GTCATCAAGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3397
QY 1141 AAGCGCTCGCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
DB 3398 GAGCGCTCGCGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3457
QY 1201 CCGTGGCGGCGGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 3458 CCGTGGCGGCGGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3517
QY 1261 GCGGCGACGAATGCGCATGCTGATTTTGGAGAGGCGCGCGCG 1302
DB 3518 GCGGCGACGAATGCGCATGCTGATTTTGGAGAGGCGCGCGCG 3559

RESULT 5

AAA92302

ID AAA92302 standard; DNA; 31422 BP.

XX AAA92302;

XX 10-JAN-2001 (first entry)

DE S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.

XX Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;

KW multifunctional enzyme; polyketide; avermectin; veterinary drug;

XX agrochemical; ds.

OS Streptomyces avermitilis.

XX Key Location/Qualifiers

FT CDS 1..14646

FT /*tag= a

PN WO200162939-A1.
 XX 30-AUG-2001.
 XX 23-FEB-2001; 2001WO-JP011381.
 PF 24-FEB-2000; 2000JP-0047405.
 XX (KYOWA) KYOWA HAKKO KOGYO KK.
 PA (KITA) KITASATO INST.
 XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 PI WPI; 2001-582053/65.
 DR P-PSDB; AAG65268.
 XX New modified avermectin aglycone synthase derived from Streptomyces
 PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX Claim 10; Page 149-167; 257pp; Japanese.
 XX The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermectilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is modified version of a
 CC fragment of the *S. avermectilis* genome.
 XX SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 99.9%; Score 1300.4; DB 22; Length 11916;
 Best Local Similarity 99.9%; Pred. No. 4.9e-214;
 Matches 1301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCCGATGCCATCATCGCATGGCATGCCGTTTCCCGGGCGAGTCCGGTCCCGAAG 60
 DB 1441 GACCCGATGCCATCATCGCATGGCATGCCGTTTCCCGGGCGAGTCCGGTCCCGAAG 1500
 QY 61 GACCTGTGGAGTGCCGCTCCGGCGGAGAGCCATCGGGCCGTTCCCAACCGACCGC 120
 DB 1501 GACCTGTGGAGTGCCGCTCCGGCGGAGAGCCATCGGGCCGTTCCCAACCGACCGC 1560
 QY 121 GGATGCCACGAAACAGCGTCAAGCCAGGACCCACCGAGCCCGGCACTTCTATCCG 180
 DB 1561 GGATGCCACGAAACAGCGTCAAGCCAGGACCCACCGAGCCCGGCACTTCTATCCG 1620
 QY 181 CAGGAGGGGGGTTCTTACAGACGGGGGCACTTCGAGCCCGGTTCTTCCGGAATCAGT 240
 DB 1621 CAGGAGGGGGGTTCTTACAGACGGGGGCACTTCGAGCCCGGTTCTTCCGGAATCAGT 1680
 QY 241 CCACGTGAGCACTGCGCATGGATCCGACAGCGGCTGCTGTGGAGACGTCCTGGAG 300
 DB 1681 CCACGTGAGCACTGCGCATGGATCCGACAGCGGCTGCTGTGGAGACGTCCTGGAG 1740
 QY 301 GCGTTGAGGGGGGGGAATCGATCCGCTGTCGGTACGGGGTCCCGTACGGGCGTCTTC 360
 DB 1741 GCGTTGAGGGGGGGGAATCGATCCGCTGTCGGTACGGGGTCCCGTACGGGCGTCTTC 1800
 QY 361 GCGGGGGCCCTCTCTTCCGACTACGGCCCGCGTATGACACCGCGTCTCGGAGGCGCC 420
 DB 1801 GCGGGGGCCCTCTCTTCCGACTACGGCCCGCGTATGACACCGCGTCTCGGAGGCGCC 1860
 QY 421 GCGGACGTGAGGGGGCACAATCTTCCACCGGTACCAACCGGCGAGCGTCTGTGGGCCGTATC 480
 DB 1861 GCGGACGTGAGGGGGCACAATCTTCCACCGGTACCAACCGGCGAGCGTCTGTGGGCCGTATC 1920
 QY 481 GCCTACAGCTTCGGGCTGGAGGGCGCGGATCACCGTGGACACGGGGGTCTCGGCATCG 540
 DB 1921 GCCTACAGCTTCGGGCTGGAGGGCGCGGATCACCGTGGACACGGGGGTCTCGGCATCG 1980

QY 541 CTCGTGACGCTGCATCTGGCGTGCAGTCTGGCTGGGTCGGGTGAGTGACGTCGCGCTG 600
 DB 1981 CTCGTGACGCTGCATCTGGCGTGCAGTCTGGCTGGGTCGGGTGAGTGACGTCGCGCTG 2040
 QY 601 GCGGCGGCGCTGCTGGTCAATGTCACCTCCGTCATGTCGATGTCATGATGTCGCGGACGCGC 660
 DB 2041 GCGGCGGCGCTGCTGGTCAATGTCACCTCCGTCATGTCGATGTCATGATGTCGCGGACGCGC 2100
 QY 661 GGGCTGTCGGTGGACGCGAGGTGCAAGCGCTACTCGGTCGACGCGAGCGGACGCGGCTGG 720
 DB 2101 GGGCTGTCGGTGGACGCGAGGTGCAAGCGCTACTCGGTCGACGCGAGCGGACGCGGCTGG 2160
 QY 721 GCGCAGGCGCTCGGATGCTGTTGGTGAGCGGTTGTGCGATGCGGTGCGGTGCGGCGAT 780
 DB 2161 GCGCAGGCGCTCGGATGCTGTTGGTGAGCGGTTGTGCGATGCGGTGCGGTGCGGCGAT 2220
 QY 781 CGGTCGTCGGCGGTGTCGACGGCGCATGTCGGTCAACAGGACGTCGCTCGAATGGGCTG 840
 DB 2221 CGGTCGTCGGCGGTGTCGACGGCGCATGTCGGTCAACAGGACGTCGCTCGAATGGGCTG 2280
 QY 841 ACGGCGCGAAGCGTCCGCTCAGGAGCGGCTGATCCGCGCAGGCGTTCGCGAAGCGCGG 900
 DB 2281 ACGGCGCGAAGCGTCCGCTCAGGAGCGGCTGATCCGCGCAGGCGTTCGCGAAGCGCGG 2340
 QY 901 TTGTCGCTGGCGGATGTCGATGTCGTCGAGGGGCAACGCGGACGCGTGGGTGAT 960
 DB 2341 TTGTCGCTGGCGGATGTCGATGTCGTCGAGGGGCAACGCGGACGCGTGGGTGAT 2400
 QY 961 CGGTCGAGGACAGGCGTTCGTCGACGTCAGGCGAGCGGCGCGGTGACAGGCGCGCTG 1020
 DB 2401 CGGTCGAGGACAGGCGTTCGTCGACGTCAGGCGAGCGGCGCGGTGACAGGCGCGCTG 2460
 QY 1021 TGGCTGGGCTCTGAACTCCAACTCCGCGCACACCATGGCTGCCGCGGCTGTCGCGTGG 1080
 DB 2461 TGGCTGGGCTCTGAACTCCAACTCCGCGCACACCATGGCTGCCGCGGCTGTCGCGTGG 2520
 QY 1081 GTCTCAAGATGCTGATGTCGCTGGGAGGGGGTGTTCGCGCGGACGTCGTCATGTCGAT 1140
 DB 2521 GTCTCAAGATGCTGATGTCGCTGGGAGGGGGTGTTCGCGCGGACGTCGTCATGTCGAT 2580
 QY 1141 AAGCGCTCCGCGCAGGTGACTGTCGCGGGGGGGTGGTCCGCTGCTGACGAGGCGGTG 1200
 DB 2581 AAGCGCTCCGCGCAGGTGACTGTCGCGGGGGGGTGGTCCGCTGCTGACGAGGCGGTG 2640
 QY 1201 CGGTGGCGGGGACGCGGAGCGGCTTCGCGGGGGGGGAGTGTGCTGTCGTCGCGATC 1260
 DB 2641 CGGTGGCGGGGACGCGGAGCGGCTTCGCGGGGGGGGAGTGTGCTGTCGTCGCGATC 2700
 QY 1261 GCGGCGACGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 1302
 DB 2701 GCGGCGACGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 2742

RESULT 4

AAZ58381
 ID AAZ58381 standard; DNA; 12381 BP.
 XX
 AC AAZ58381;
 XX
 DT 23-MAY-2000 (first entry)
 XX
 DE Streptomyces avermectilis avermectin polyketide synthase modules 1+2.
 XX
 KW Polyketide synthase; avermectin; insecticide; ss.
 OS Streptomyces avermectilis.
 XX
 PN WO200001827-A2.
 XX
 PD 13-JAN-2000.
 XX
 PF 06-JUL-1999; 99WO-GB02158.

PD 30-AUG-2001.
XX
XX PF 23-FEB-2001; 2001WO-JP01381.
XX PR 24-FEB-2000; 2000JP-0047405.
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI; 2001-582053/65.
DR P-PSDB; AAG65264, AAG65265.
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
XX Example 2; Page 58-123; 257pp; Japanese.
XX
XX The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 1302; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred No. 2.5e-214;
Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GACCCGATCGCCATCATCGGCATGCGATCCGTTTCCGGGCGGAGTCGGTCCCGAAG 60
1441 GACCCGATCGCCATCATCGGCATGCGATCCGTTTCCGGGCGGAGTCGGTCCCGAAG 1500
61 GACCTGTGGAGCTGGCCCTCGGGCGGAGACGCCATCGGCGCTTCCCAACCGACGCG 120
1501 GACCTGTGGAGCTGGCCCTCGGGCGGAGACGCCATCGGCGCTTCCCAACCGACGCG 1560
121 GGATGGCCCAAGAGGTCAGCCAGGACCCAGCCAGCCGCGACGTTCTATCCG 180
1561 GGATGGCCCAAGAGGTCAGCCAGGACCCAGCCAGCCGCGACGTTCTATCCG 1620
181 CAGGAGGCGGTTCTTTCAGACGCGGCGCACTTTCAGACGCGGCTTCTTCGGAATCAGT 240
1621 CAGGAGGCGGTTCTTTCAGACGCGGCGCACTTTCAGACGCGGCTTCTTCGGAATCAGT 1680
241 CACGTGAGGCACTGGCGATGATCCGACGAGCGGCTGCTGAGACGTCCTGGAG 300
1681 CACGTGAGGCACTGGCGATGATCCGACGAGCGGCTGCTGAGACGTCCTGGAG 1740
301 GCGTTGAGCGCGGGAATCGATCCGCTGCGGTACCGGCTCCGTCACGGCGCTTC 360
1741 GCGTTGAGCGCGGGAATCGATCCGCTGCGGTACCGGCTCCGTCACGGCGCTTC 1800
361 GCGGGCGCCCTCTCTTCGACTACGCGCCGCGTATGGACACCGCGTCGTTCGAGGCGCC 420
1801 GCGGGCGCCCTCTCTTCGACTACGCGCCGCGTATGGACACCGCGTCGTTCGAGGCGCC 1860
421 GCGGACGTGGAGGCGCAATCTTACCGGTACACGCGGAGCGTCCTGTTCGGGCGGTATC 480
1861 GCGGACGTGGAGGCGCAATCTTACCGGTACACGCGGAGCGTCCTGTTCGGGCGGTATC 1920
481 GCTACAGCTTCGGGCTGGAAGGCGCGGATCACCGTGGACACGCGGTCTCGGCATCG 540
1921 GCTACAGCTTCGGGCTGGAAGGCGCGGATCACCGTGGACACGCGGTCTCGGCATCG 1980
541 CTCGTGACGCTGCATCTGGCGTGCAGTCTGCTGCGGTGAGTGCACGCTCGCGCTG 600

Db 1981 CTCGTGACGCTGCATCTGGCGTGCAGTCTGCGGTGAGTGCACGCTCGCGCTG 2040
Qy 601 GCGGGCGGCTCTCGGTCAATGTCACCTTCGSCATGTTTCATCGAGTTCCTCCGCGAGCGC 660
Db 2041 GCGGGCGGCTCTCGGTCAATGTCACCTTCGSCATGTTTCATCGAGTTCCTCCGCGAGCGC 2100
Qy 661 GCGGTCTCGGTGACGCGGAGTGCAGGCGTACTCGGTGACGCGGAGTGCAGGCGGCTG 720
Db 2101 GCGGTCTCGGTGACGCGGAGTGCAGGCGTACTCGGTGACGCGGAGTGCAGGCGGCTG 2160
Qy 721 GCGGAGGCGCTCGGATGCTGTTGAGCGGTTGTCGATGCGGTGCGGTGCGGTGCGGCTAT 780
Db 2161 GCGGAGGCGCTCGGATGCTGTTGAGCGGTTGTCGATGCGGTGCGGTGCGGTGCGGCTAT 2220
Qy 781 GCGGTCTCGGTGAGTGCAGGCGGAGTGCAGGCGTCAACAGGAGCGGTGCGGTGCGGTG 840
Db 2221 GCGGTCTCGGTGAGTGCAGGCGGAGTGCAGGCGTCAACAGGAGCGGTGCGGTGCGGTG 2280
Qy 841 ACGGCGCGAAGCGGTCCGGCTCAGGAGCGGCTGATCCGGCGAGCGGTTCGCGAAGCGCGGG 900
Db 2281 ACGGCGCGAAGCGGTCCGGCTCAGGAGCGGCTGATCCGGCGAGCGGTTCGCGAAGCGCGGG 2340
Qy 901 TTGTCGCTGGCGGATGTCGATGTCGAGGCGGACCGGACCGGACCGGACCGGTCGTCGAT 960
Db 2341 TTGTCGCTGGCGGATGTCGATGTCGAGGCGGACCGGACCGGACCGGACCGGTCGTCGAT 2400
Qy 961 CCGATCGAGGCGACAGGCGCTGCTGCCACGTCACGCGGAGCGGCGGTCGACAGGCGCGCTG 1020
Db 2401 CCGATCGAGGCGACAGGCGCTGCTGCCACGTCACGCGGAGCGGCGGTCGACAGGCGCGCTG 2460
Qy 1021 TGGCTGGGCTCTTGAAGTCCAAACATCCGGCGACACCATGCGTTCGCGCGGCTGTCGTCGG 1080
Db 2461 TGGCTGGGCTCTTGAAGTCCAAACATCCGGCGACACCATGCGTTCGCGCGGCTGTCGTCGG 2520
Qy 1081 GTCATCAAGATGTCGATGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGAT 1140
Db 2521 GTCATCAAGATGTCGATGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGAT 2580
Qy 1141 AAGCGCTCGCGGACGTCGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGAT 1200
Db 2581 AAGCGCTCGCGGACGTCGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGAT 2640
Qy 1201 CCGTGGCGGCGGACGTCGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGATC 1260
Db 2641 CCGTGGCGGCGGACGTCGTCGTCGAGGCGGTCGTCGCGGACGTCGTCGTCGTCGTCGATC 2700
Qy 1261 GCGGCGACGAATGCGCATGTCATTTTGAGGAGGCGCGCGG 1302
Db 2701 GCGGCGACGAATGCGCATGTCATTTTGAGGAGGCGCGCGG 2742
RESULT 3
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
XX AC AAH79279;
XX
XX DT 04-DEC-2001 (first entry)
XX
XX DE Streptomyces avermectilis coding sequence derivative SEQ ID NO: 3.
XX
XX KW Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide; ds.
XX
XX OS Synthetic.
XX
XX FH Key
XX CDS
XX
XX FT 1. 11916
XX /tag= a
XX /product= "AAG65268"
XX /partial
XX

(KITA) KITASATO INST.
 Omura S, Ikeda H;
 MPI; 2000-565458/52.
 DR P-PSDB; AAB23749, AAB23750.
 XX
 Avermectin aglycone synthase DNA and proteins encoded by all or part of
 it for the production of avermectin and its derivatives for drug and
 agrochemical use -
 Claim 2; Page 66-134; 314pp; Japanese.
 CC The present sequence represents DNA which encodes avermectin aglycon
 synthase proteins. Also described are: (1) polypeptides encoded by all
 or part of the DNA; (2) expression vectors containing the DNA; (3) host
 cells transformed by the vectors; (4) preparation of the polypeptides
 by culture of the transformants; (5) preparation of avermectin aglycon
 or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 containing sense or antisense sequences from the avermectin aglycon
 synthase DNA. The enzymes are useful for the production of modified
 CC forms of avermectin and of the intermediates in its biosynthesis, for
 CC use as drugs, veterinary drugs and agrochemicals.
 XX
 SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;
 Query Match 100.0%; Score 1302; DB 21; Length 30690;
 Best Local Similarity 100.0%; Pred. No. 2.5e-214;
 Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 P-PSDB; AAB23749, AAB23750.
 1441 GACCCGATCGCCATCATCGCATGCGATGCGTTTCCCGGGCGAGTCCGGTCCCGAAG 60
 1441 GACCCGATCGCCATCATCGCATGCGATGCGTTTCCCGGGCGAGTCCGGTCCCGAAG 1500
 61 GACCTGTGGAGTGGCGCTCGGGCGAGAGCGCATCGGGCGTTTCCCAACCGACCGC 120
 1501 GACCTGTGGAGTGGCGCTCGGGCGAGAGCGCATCGGGCGTTTCCCAACCGACCGC 1560
 121 GGATGCCCCACGAGCGTACGCCCCAGGAGCCCGACCGCCCGGCGTCTTATCCG 180
 1561 GGATGCCCCACGAGCGTACGCCCCAGGAGCCCGACCGCCCGGCGTCTTATCCG 1620
 181 CAGGAGGGCGGTTCCTTTCACAGCGGGCGACTTCGACGCGGCTTTCGGAATCAGT 240
 1621 CAGGAGGGCGGTTCCTTTCACAGCGGGCGACTTCGACGCGGCTTTCGGAATCAGT 1680
 241 CCACGTGAGGCACTGGCGATGATCCGACGAGCGGCTGCTGGAGAGCTCTGGGAG 300
 1681 CCACGTGAGGCACTGGCGATGATCCGACGAGCGGCTGCTGGAGAGCTCTGGGAG 1740
 301 GCGTTGAGCGGGCGGGAATCGATCCGTTGCGTACGCGGGTCCCGTACGGGCGTCTTC 360
 1741 GCGTTGAGCGGGCGGGAATCGATCCGTTGCGTACGCGGGTCCCGTACGGGCGTCTTC 1800
 361 GCGGGCGGCTCTCTTCGATACGCGCGCGTATGACACCGCGTCTCGAGGGCGCC 420
 1801 GCGGGCGGCTCTCTTCGATACGCGCGCGTATGACACCGCGTCTCGAGGGCGCC 1860
 421 GCGGAGTGGAGGGCCACATCTTACCGGTACACCGGAGCGTCTGTCGGGCGGTATC 480
 1861 GCGGAGTGGAGGGCCACATCTTACCGGTACACCGGAGCGTCTGTCGGGCGGTATC 1920
 481 GCCTACAGCTTCGGGCTGGAGGGCGGCGATCACCCTGGACACGGGGTCTCGGCATCG 540
 1921 GCCTACAGCTTCGGGCTGGAGGGCGGCGATCACCCTGGACACGGGGTCTCGGCATCG 1980
 541 CTGCTGACGCTGATCTGGGCTGCCAGTCTCGGTCGGGTGAGTGACAGCTCGGCTG 600
 1981 CTGCTGACGCTGATCTGGGCTGCCAGTCTCGGTCGGGTGAGTGACAGCTCGGCTG 2040
 601 GCGGCGGCGTCTCGGTCATATGTCACCGCTCGGCATCTTATCGAGTTCTCCGGCAGCGC 660

Db 2041 GCCGCGGCGGTCTCGGTCAATGCCACCTCCGATGTTTCATCGATTTCTCCGCGCAGCGC 2100
 QY 661 GGGCTGTGCTGGAGCGGCGAGGTGCAAGCGGTACTCGGCTGCGAGCGGCGACCGGCTGG 720
 Db 2101 GGGCTGTGCTGGAGCGGCGAGGTGCAAGCGGTACTCGGCTGCGAGCGGCGACCGGCTGG 2160
 QY 721 GGCAGGCGGTGGGATGCTGTTGGTGAGCGGTTGTGCGATGCGGTGCGGCTGGGCGAT 780
 Db 2161 GGCAGGCGGTGGGATGCTGTTGGTGAGCGGTTGTGCGATGCGGTGCGGCTGGGCGAT 2220
 QY 781 GCGGTGCTGGGCGGTGATACGCGGCGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 840
 Db 2221 GCGGTGCTGGGCGGTGATACGCGGCGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 2280
 QY 841 ACGGCGCGAAGCGGTGCGGCTCAGGAGCGGCTGATCCGGCAGGCGTTCGCGAAGCGGCG 900
 Db 2281 ACGGCGCGAAGCGGTGCGGCTCAGGAGCGGCTGATCCGGCAGGCGTTCGCGAAGCGGCG 2340
 QY 901 TTGTCCTGGGCGGATGTTGGATGTTGGTGAGGGGCGACGGGCGACGACGCTGGGTGAT 960
 Db 2341 TTGTCCTGGGCGGATGTTGGATGTTGGTGAGGGGCGACGGGCGACGACGCTGGGTGAT 2400
 QY 961 CCGATCGAGGCGACAGGCGTTCGTCGACGATCGGCGAGCGGCGGCTGACAGGCGGCTG 1020
 Db 2401 CCGATCGAGGCGACAGGCGTTCGTCGACGATCGGCGAGCGGCGGCTGACAGGCGGCTG 2460
 QY 1021 TGGCTGGGCGTCTCGAGTCCAAACATCGGCGACACCATGCGCGGCGGTGTTGGGTGG 1080
 Db 2461 TGGCTGGGCGTCTCGAGTCCAAACATCGGCGACACCATGCGCGGCGGTGTTGGGTGG 2520
 QY 1081 GTCAATCAAGATGTTGATGCGGTTCGCGGAGGGGCGTTCGCGCGGCGACGTTGCATGTG 1140
 Db 2521 GTCAATCAAGATGTTGATGCGGTTCGCGGAGGGGCGTTCGCGCGGCGACGTTGCATGTG 2580
 QY 1141 AAGCGTCCGCGAGGTGACTGTTGCTCCGCGGGGCGGTCGCGGCTGCTACGAGGCGGCTG 1200
 Db 2581 AAGCGTCCGCGAGGTGACTGTTGCTCCGCGGGGCGGTCGCGGCTGCTACGAGGCGGCTG 2640
 QY 1201 CCGTGGCGGGGCGGCGGCGGCGGCTTCGCGGGGCGGAGTGTGCTGTTCCGGATC 1260
 Db 2641 CCGTGGCGGGGCGGCGGCGGCGGCTTCGCGGGGCGGAGTGTGCTGTTCCGGATC 2700
 QY 1261 GCGGCGACGAATGCGCATGTGATTTTGAGGAGGCGCGCGCG 1302
 Db 2701 GCGGCGACGAATGCGCATGTGATTTTGAGGAGGCGCGCGCG 2742
 RESULT 2
 AAH79277
 ID AAH79277 standard; DNA; 30690 BP.
 XX
 AC AAH79277;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Streptomycetes avermectin coding sequences SEQ ID NO: 1.
 XX
 KW Avermectin aglycone synthase; AAS; avermectin derivative;
 KW drug production; veterinary drug; pesticide; ds.
 XX
 OS Streptomycetes avermectin.
 XX
 FH Key Location/Qualifiers
 CDS 1..11919
 FT /*tag= a
 FT /product= "AAG65264"
 FT /partial
 FT CDS 11971..30690
 FT /*tag= b
 FT /product= "AAG65265"
 XX
 PN WO200162939-A1.
 XX

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1302	100.0	30690	21	AAA92301	<i>S. avermitilis</i> ave
2	1302	100.0	30690	22	AAH79277	<i>Streptomyces</i> averm
3	1300.4	99.9	11916	21	AAH79279	<i>Streptomyces</i> averm
4	1298.8	98.8	12381	21	AAZ58381	<i>Streptomyces</i> averm
5	944.8	72.6	31422	21	AAH52302	<i>S. avermitilis</i> ave
6	944.8	72.6	31422	22	AAH79278	<i>Streptomyces</i> averm
7	710	54.5	20394	22	AAAF24892	Pimaricin biosynth
8	704.8	54.1	13987	18	AAH90415	Hybrid swmG/tylG O
9	704.8	54.1	44377	18	AAT78508	Platenolide syntha

ALIGNMENTS

XX

XX
XX
REF ID: A66907
FEB-17 2016
100-408606

Db 94523 CTGTGGCTGGGTGGGTGAAGTCCAACTGGGCGCACACGACGCGCGCGCGCGCG 94464
Qy 1078 GGGGTATCAAGATGATGGTGGTGGGAGGGGGTGTGGCGGAGCGTTGCATGTG 1137
Db 94463 GGATCATCAAGATGATGATGGTATGCGGTATCGGGTGTGCGGAGGACGTCGATGTG 94404
Qy 1138 GATAAGCCGTGCGCGCAGGTGGAATGCTCGCGGGGGCGGTGCGGTGCTGACGAGGGCG 1197
Db 94403 GACCGCCGTGCGCGCAGGTGGAATGCTCGCGGGGCGCGGTGCTGACGAGGAG 94344
Qy 1198 GTGCGGTGCGCGGGGAGCGCGGAGGGGTGGTGGCGGGCGGGAGTGTGCTGTTGGG 1257
Db 94343 CGGGAGTGGCGG-----GACGCGGGCGGTGCGCGCGGGCGCGCGGTGCTGTTGGG 94290
Qy 1258 ATCGCGGCGCAGGATCGCATGTGATTTTGGAGGCGCGCGCG 1301
Db 94289 ATCAGCGGCGCAGGATCGCATGTGATTTTGGAGGCGCGCGCG 94246
RESULT 15
AX250262
LOCUS AX250262 30000 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 2 from Patent WO0168867.
ACCESSION AX250262
VERSION AX250262.1 GI:15984064
KEYWORDS Streptomyces cinnamonensis.
SOURCE Streptomyces cinnamonensis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30000)
AUTHORS Leadlay, P.F., Staunton, J. and Ollinyk, M.C.
TITLE Polyketides and their synthesis
JOURNAL Patent: WO 0168867-A 2 20-SEP-2001;
Biotech Technology Limited (GB)
FEATURES
source
1..30000
/organism="Streptomyces cinnamonensis"
/db_xref="taxon:1900"
/note="Nucleotides 30001 - 60000 of the monensin
biosynthetic gene cluster"
BASE COUNT 3911 a 10993 c 10862 g 4234 t
ORIGIN
Query Match 52.0%; Score 677; DB 6; Length 30000;
Best Local Similarity 71.8%; Pred. No. 1.1e-76;
Matches 935; Conservative 0; Mismatches 350; Indels 18; Gaps 3;
Qy 1 1 GACCCGATCGCCATCATCGCATGGCATGCGGTTTCCCGGGCGGAGTCCGGTCCCGGAAG 60
Db 17944 GACCCGATCGCCATCATCGCATGGCATGCGGTTTCCCGGGCGGAGTCCGGTCCCGGAC 18003
Qy 61 GACCTGTGGAGTGGCGCTCGGCGGAGAGCGCATCGGGCGGTTCCTCCACCGACCGC 120
Db 18004 GACCTGTGGAGTGGCGCTCGGCGGAGAGCGCATCGGGCGGTTCCTCCACCGACCGC 18063
Qy 121 GGATGCCACCGAAGAGGTGACGCGGAGAGCGCGGAGCGCGGCGGTTCCTCCACCGACCGC 180
Db 18064 GGCTGGGACCTGGAAAGCGCTTTCACCGCGGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 18123
Qy 181 CAGGAGAGCGGGTTCCTTCAACGCGGGCGGCGCTTCGAGCGCGGCGGTTCCTCGGAATCAGT 240
Db 18124 GACGAGGCGGATTTCTTCCGAGCGGGGTGATTTGATGGCGGCTTCTTCGGGATCAT 18183
Qy 241 CCACGTGAGCATGCGGATGATGATGCGAGAGCGGCTGCTGTGAGAGCGTCTGGGAG 300
Db 18184 CCGCGGAGGCGCTGCGGATGATGATGCGAGAGCGGCTGCTGTGAGAGCGTCTGGGAG 18243
Qy 301 GCGTTCAGCGGCGGAGATCGATCGGCTGTGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db 18244 GTGTTGGAGCGTGGGGTATCGATCGAGCGCTCAAGGGGACCGCGGACCGGACCTAC 18303

Qy 361 GCGGCGCGCTCTCTCTTCGATACGCGCGCGGTATGGAACACGCGTCTGTCGAGGGCGCC 420
Db 18304 GTGGGCTCATGTATACACGACTACGCCAAGTCTTCCACACGCG-----CGAC 18351
Qy 421 CGGAGCGTGGAGGGCGCATCTCTACCGGTACACGCGGACGCGTCTCTGTCGGGCGGTATC 480
Db 18352 GCGGAGTGGAGGGCTACTCTCTTGTGCGAGCACCGGACGATGGTCTCCGCGCGCGTTC 18411
Qy 481 GCCTACAGCTTTCGGGCTGGAAGGGCGGCGATACCGTGGACACGGGGTGTCTCGGCATCG 540
Db 18412 GCCTACACCTTGGGCTTGAAGGTTCGGGCGGTGAGGTGAGCGTGCACACCGCGTGTCTCTCTCC 18471
Qy 541 CTCTGAGCGCTGCAATCTGGCGTGCAGTTCGTGCGGTGCGGTGAGTGCACGCTCGCGCTG 600
Db 18472 CTGGTCTCATCTCCACCTGGCGACGCGAGCACTCCGCGACGCGGAGTGCAGACTTCGCGCTG 18531
Qy 601 GCGGCGCGCTCTCGGTCTATGTCCACCTCGGCATGTTTCATCGAGTTCCTCCCGGCGCGC 660
Db 18532 GCAGCGGTGTGACCGTCTATGCGCGACCCCGGACATGTTTCGCGGGCTTCTCGCGCGCGC 18591
Qy 661 GGGCTGTCTGGTGGAGCGGCGGTGCAAGCGTACTCTCGGCTGCAGCGCGCACCGGCTGG 720
Db 18592 GGCCTCTCACCTGACGGCGGTGCAAGGCTTACGCGCGCGGCGCGAGTGCAGTTCG 18651
Qy 721 GCGAGGCGCTCGGATGCTGTGTGGAGCGGTTGTGCGATGCGGTGCGGTGCGGTGCGGTG 780
Db 18652 TCCGAGGCGAGTGGCGGTATTGCTCTTCTGAGCGGTGTCGATGCGCGCGTCTATGGCGT 18711
Qy 781 CGGTGTCTGGCGGTGTGACGGCGCGGTGCGGTGCAACGAGGAGCGTTCGTCGATGCGGTG 840
Db 18712 CGGTGTGCGGTGTGTCGCGGTTCGCGCGGTGTAATCAGGACGCGTTCGAGTAAATGGGTG 18771
Qy 841 ACGGCGCGAAGCGTCCGCTCAGAGCGGCTGATCCGCGAGGCGTTCGCGAAGCGCGG 900
Db 18772 ACGGCGCGAAGTGTTCGTCGAGGAGCGGCTGATTCGTCAGGCGTTCGCGAGTGGTGG 18831
Qy 901 TTGTCGTGGCGGATGTGATGTGTGAGGGGCGACGCGGCGCACGACGCTGGGTGAT 960
Db 18832 TTGTCGTGGTGTGATGATGTGTGAGGGCGATGCGGACCGGAGCACGTTGGGTGAT 18891
Qy 961 CGATCGAGGCGACAGGCGGTGCTCGCCACGTTACGGGCA---CGCGGCGGTGACAGGCG 1017
Db 18892 CCGATCGAGGCGCAGGCTCTGCTGCGCCACATATGGGCGAGGGCGCTCCGAGGACCGTCCG 18951
Qy 1018 CTGTGGCTGGGTCTCTCAAGTCCAAATCGGGCACACATGGCTGCGCGCGGTGTGGGT 1077
Db 18952 TTGTGTTGGGTGCGTGAAGTGCNAATTTGTCATACGAGGCGGCTGCGGCTGTGG 19011
Qy 1078 GGGGTCAACAAGATGATGCGGTTCGCGGAGGGGGTGTTCGCGCGGACGTTGCATGTG 1137
Db 19012 GGTGTCAACAAGATGATGCGGTTCGCGGATGCGGATGCGGTGCGGCGGAGTTTGCATGTG 19071
Qy 1138 GATAAGCGCTGCGCGCAGGTGCACTGTGTCGCGGGGGCGGTGCGGCTGCTGACGAGGCG 1197
Db 19072 GATGTGCGCTGCGCGCATGTGAGTGGGATTCGCGGTGCGGTGCGGTGCGGTGAGTTCG 19131
Qy 1198 GTGCGCTGCGCGGGGAGCGCGAGCGGCGGTTCGCGCGGGCGGAGTGTGCTGCTTCCGG 1257
Db 19132 GTGCCATGGCC---GCAGGTGAGGCTGCTCGCGCTCGGGGGGTGTGTGCTGCTTCCG 19188
Qy 1258 ATCGCGGCGCAAGATGCGCATGTGATTTTGGAGGAGGCGCGCG 1300
Db 19189 GCTTCGGGCGCAATGCGCACGTCATCTGGAGTCTGTTCGCG 19231

Search completed: June 17, 2003, 17:55:28
Job time : 2533.07 sec

ASLVLAQALVDLGRVGGPRLMLVTRGAVVAGPSDAGVVIDPVQAVQVWGFRVLGLEH
PRLMGLVDPVGVDEBVCRRFVGVASAGFEDQVARGSGVWVRRLVRAVVDGGCGG
WPRGTVLVGLGLGAGTARWLKGGGADHVIVSRGSGAPAGDVLHLEGLGGA
RVSACDADVADVALRALLSDIGEPVAVFHAAGVQPSQPLAEISVQEAADVMAKVA
GAVNUGELVPCGLEAFVLFSSNAGVGGGQAVYAANAFDIALAARRRGGVGLPATS
VANGWAGMGSAVGAARELSRRGRVAMRPERAVAMADAGEAFVAVADVDMER
FVTGASARPLRILSDLEPRVAVVEGQVGGGLGVEEESGWLKLSLSRVRO
EEELVELVRAQAAVILGHGSAQDVPAERAFKELGDFSLTAVELRNLAAATGIRLPAT
MAFDHPTAIAIRFLQSELVGGDDPLTLMRSAIDOLETGLALLEDEEARSSEITKRLN
IILPRGSGSSRGRAGDAGEHQDVEATIDELFEVLDELNGNS"
630..1906
/genes="oleAI"
/notes="Region: ketoacyl synthase domain KS-Q"
2309..3415
/genes="oleAI"
/notes="Region: acyltransferase domain AT-L"
3485..3724
/genes="oleAI"
/notes="Region: acyl carrier protein domain ACP-L"
3806..5056
/genes="oleAI"
/notes="Region: ketoacyl synthase domain KS1"
5363..6400
/genes="oleAI"
/notes="Region: acyltransferase domain AT1"
7109..7669
/genes="oleAI"
/notes="Region: ketoacyl reductase domain KR1"
7943..8200
/genes="oleAI"
/notes="Region: acyl carrier protein domain ACP1"
8258..9541
/genes="oleAI"
/notes="Region: ketoacyl synthase domain KS2"
9848..10882
/genes="oleAI"
/notes="Region: acyltransferase domain AT2"
11723..12274
/genes="oleAI"
/notes="Region: ketoacyl reductase domain KR2"
12560..12817
/genes="oleAI"
/notes="Region: acyl carrier protein domain ACP2"
13118..24568
/genes="oleAI"
13118..24568
/genes="oleAI"
/notes="polyketide synthase"
/codon_start=1
/transl_table=11
/product="8_8a-deoxyoleandolide synthase 2"
/protein_id="AAF82409.1"
/db_xref="GI:9049536"
/translation="MTNDEKIVEYLRKATVDLIRKARHRIWELEDEPIATTSMACHFPG
GIESPOLWELLSAGVEISSEFPDDRDNDLDEIYHPDEHSGTSYVRHGGFLDHATQF
DIDFGISPREALAMDPOORLLLETSMQWLFERAGVDHPHTLKGRSGVFVGAAMGVAD
RVDPPEAEVGLLTGNASAVSGRISYTFGLEGPVAVTDACSSLSLHVALHVAOLR
RCECSLAVGVGVAWSDPKVFVEFSRQGLARDGRSKAFASADGFGFAEGVSLILLE
RUSLARLRVLVAVRGVAVNQDASNGLAAPGPSQQRVIRALADAGLAPADV
VBAHGTVRLGDDPIEAQALLATYQGGTSGRPVWLGSVKSNTGHTQAAGVAGVWKMV
LALERGVPKTLHVDPSPHVMSGTAVELLTEERPEPEABRLRAGISAFVSGSTN
AHVIVPEAPAEPEPEPGTRVVAAGDLVVPVWVSGRDAGALRAQARLAHVSTGA
GVVDVCMVSLVATRSVFEHRAVMVQTDLSMAGSLAGFAAGVVPVGVSVAPAEGRV
VVPVPGGQSMVGMAGLLDACPVFAEVAECAVLDPLTCHSLVEVLGGEAIVGRV
DVQPALWAVMSLARTWRYGYVEPAAVVGHSGEIAAACVAGGLSLADGARVVILRS
RAIARTAGGGVWVSLPAGRVTRMLDITYGRLSVAAVNGPSSTVSGDAQLDELLA
GCEGREGVARRVPVDVYASHAOMDQLRDELLEALADITPQHSVPFFSTVTDWLDTT
ALDAGVFTNLRATVPEAVGLVAQGMGAFVCEGSPHVPVPGIEOTLDTVADAVA
LGSRLDEGLGRFLTSLAEAFVQGVVPVMSRTFEGASPRVDLPTYPFORORFWLEG
SPALSNVGEEDAVPMDAVEREDSAVABELGIDAKALHMTLPAISWRRRQRQR
KVQWRYRVEWKRPLNSAGSLLVVPQGRAGDVRTQSVAVAAKGEATVL
EVDALHPDRAAVAEALTRWPGVGVVSVLAWEEQALAEHPVLSAGLAASLALAQLID
VGGSGESAPRLMLVTEAAVVI GAADTGAVIDPVHQAOLWGFGRVLAEHPELMGLLID

Query Match 52.9%; Score 689; DB 1; Length 24568;
Best Local Similarity 72.3%; Pred. No. 3.5e-78;
Matches 944; Conservative 0; Mismatches 340; Indels 21; Gaps 3;
QY 1 GACCGATCGCATCATCGGATCGCATCGCGTTCCTCCGGCGGAGTCCGGTCCCGGAG 60
DB 8258 GAGCGATCGCGATCGTGGGAATGGCATGCGCGTTCCTCCGGCGGAGTCCGGTCCCGGAG 8317
QY 61 GACCTGTGGGAGCTGGCGGCTCGGGCGGAGACGCGATCGGCGGCTTCCCGACCGACCGC 120
DB 8318 GATCTGTGGAGCTGCTCACCCTGGTGAGGACGTTGTCTCCGACTCCCGCAGAACCGC 8377
QY 121 GGATGGCCCGGAAACAGCGTCACGCCCGAGACCCACGACGCCCGGACGTTTATCG 180
DB 8378 GGGTGGGACTTGGGCGGCTGTACGATCCGATCCGAGACCGGCGGCTACGTCATACATG 8437
QY 181 CAGGAGCGGGTTCCTTTCAGAGCGGCGGCGACTTCGAGCGCGGCTTCTTCGGGATCATG 240
DB 8438 COTGAGGGTGTCTTCTCTGACAGCGGGGGAGTTCGAGCGGCGCTTCTTCGGTATCTCG 8497
QY 241 CCACGTGAGGCATCGCGATCGGATCGCGAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
DB 8498 CCGCTGAGGCGTGGCGATGGACCCCGCAGCAGCGGTTGCTGCTGGAGACGTCCTGGGAG 8557
QY 301 GCGTTCGAGCGGCGGGAATCGATCCGCTGCTGCTGCGGCTCCCGTACGGGCGTCTTC 360
DB 8558 GCGCTCGAAACGGCGCGCATCACTCTTCCAAGCTGGCGGGCAGTCCGACCGGTGTTC 8617
QY 361 GCGGCGGCGCTCTCTTCGACTACGGCCCGCGGTATGGACACCGCGCTCGTCGGAGGCGCC 420
DB 8618 TTCGCGCATGTGAAACACGAGACTACGCCGCCCGAGGGCGGCGAGTCCGCTCGCG 8671
QY 421 GCGGACGTGGAGGGCCACATCTCTACCGGTACCAACCGGCGAGCGTCTCTCGGCGCGTATC 480
DB 8672 -----CTGAGGGCTACTGCTCTACCGGCTCCATCTCCAGCGTCTCGGCGGTGT 8725
QY 481 GCCTACAGCTTCGGGCTGGAGGGCGCGGATACCGGTGGACACGGGGTGTCTCGGCATCG 540
DB 8726 GCTTACAGCTTCGGTCTTGGAGGGCGCTCGGCTGACGGTGGATACCGCGTGTTCGCTCG 8785
QY 541 CTCGTGACGCTGCTCTCGGCTGCGAGTCTCGCTCGGCTGAGTGCACGCTCCGCGCTG 600
DB 8786 TTGGTGGCGTGTGATCTCGCGGTGCGAGGGGTTGCGCGGGGTGAGTGTTCGCTTCGCTG 8845

QY	661	GGGCTGTCGGTGAACGGCAGGTGCAAGGGCGTACTCGGCTGCAGCCGACCGCCGCTGG	720
Db	70616	GGACTGTCGGCGACCGCCGCTGCTCAAGGGCGTTGCGCGATGCGCGCCGACGCGCTCGGCTGG	70557
QY	721	GGCGAGGGCGTGGGATGCTGTTGGTGGAGCGGTGTCGGATGCGGTGCGGCTGGGGCAT	780
Db	70556	GGCGAGGGCGTGGGATGCTGTTGGTGGAGCGGTGTCGGATGCGGTGCGGCTGGGGCAT	70497
QY	781	CGGGTCTGTCGGGTGATGCGGCGAGTGGCGGTCAACACGAGGACGGTGCATGCGATGGGCTG	840
Db	70496	CGGGTCTGTCGGGTGATGCGGCGAGTGGCGGTCAACACGAGGACGGTGCATGCGATGGGCTG	70437
QY	841	ACGGCCCGCAACGGTCCGGCTCAGAGCGGGTATCCGGACGCGTGTGGCGAACCGCGGG	900
Db	70436	ACGGCCCGCAACGGTCCGGCTCAGAGCGGGTATCCGGACGCGTGTGGCGAACCGCGGG	70377
QY	901	TTGTCGTCGGGATGCGATGCGGCGGTGGAGGGGCGACGGGACGCGGACGCGTGGGTGAT	960
Db	70376	TTGTCGTCGGGATGCGATGCGGCGGTGGAGGGGCGACGGGACGCGGACGCGTGGGTGAT	70317
QY	961	CCGATCGAGGCAACGCGTTCGTCGCGACGTCACGG---CGACGGGCGCGTGCACGCGCG	1017
Db	70316	CCGATCGAGGCAACGCGTTCGTCGCGACGTCACGG---CGACGGGCGCGTGCACGCGCG	70257
QY	1018	CTGTGCTGGGTCTCTGAAGTCCAAATCGGCGACACCATGCTGCGCGCGGTGTGGGT	1077
Db	70256	CTGTGCTGGGTCTCTGAAGTCCAAATCGGCGACACCATGCTGCGCGCGGTGTGGGT	70197
QY	1078	GGGTCATCAAGATGCGATGCGGTGCGGAGGGGGTGTGCGCGGCGCGTGCATGCTG	1137
Db	70196	GGTGTCAATCAAGATGCGATGCGGTGCGGAGGGGGTGTGCGCGGCGCGTGCATGCTG	70137
QY	1138	GATAAGCGCTGCGCGCAGTGCAGTGGTCCGCGGGGGCGTGCCTGCTGACGAGGCG	1197
Db	70136	GATGAACCGTCTGCGATGTCGATGAGTGCCTGCGGCTGCGAGTGCCTGACCTCCGAG	70077
QY	1198	GTGCGCGTGGCGGGGACCGCGAGGCGGTTGCGCGCGCGCGGAGTGTGCTGCTCGG	1257
Db	70076	GCGGAGTGGCGCGAGGCGGAGG-----GCGCGCGCGCGCGGCGTCTCCTCTCGGC	70023
QY	1258	ATCGCGCGGACCAATGCGCATGATTTGGAGGAGCC	1295
Db	70022	GTGACGCGGACCAATGCGCATGATTTGGAGGAGCC	69985
RESULT	12		
AF220951			
LOCUS	24568 bp	DNA	linear
DEFINITION	Streptomyces antibioticus 8,8a-deoxyoleandolide synthase 1 (oleA1) and 8,8a-deoxyoleandolide synthase 2 (oleA2) genes, complete cds.		
ACCESSION	AF220951		
VERSION	AF220951.1	GI:9049534	
KEYWORDS	Streptomyces antibioticus.		
SOURCE	Streptomyces antibioticus		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.		
REFERENCE	1 (bases 1 to 24568)		
AUTHORS	Shah, S., Xue, Q., Tang, L., Carney, J.R., Betlach, M. and McDaniel, R.		
TITLE	Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin		
JOURNAL	J. Antibiot. 53 (5), 502-508 (2000)		
MEDLINE	20363406		
PUBMED	10908114		
REFERENCE	2 (bases 1 to 24568)		
AUTHORS	Shah, S., Betlach, M. and McDaniel, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JAN-2000) Korean Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA		
FEATURES	Location/Qualifiers		
source	1..24568		
	/organism="Streptomyces antibioticus"		

gene
CDS

/db_xref="taxon:1890"
623..13075
/gene="oleA1"
623..13075
/gene="oleA1"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="8,8a-deoxyoleandolide synthase 1"
/protein_id="AAF82408.1"
/db_xref="GI:9049535"

/translation="MHVPEENHSHIALVIGIACRLPGSATPQEFWRLLADSDADLDEP
PAGRPPTGSLSPAPPCGFLDSIDTFADFNISPREAGVLDPOORLALELGLWALE
DAGVPHRLRGTTSVFMAWMDYAHILAHGEALTRHSLTGTTHGMIANRLSYAL
GLQPSLTVDTGQSSSLAAVHMACESLARGESDLAVGVNLTDPAGTGTGVRGAL
SPDRCYTFDSRANGYRGEQVVLKPTHRALADGTVCEILGSNNDNATGEG
TVPSRAQDVILRQWENARVAPTQVVLHGTGTPAGDPVEAGLGTALGTARPAE
APLVGSKVTHIGHLEAGIAGLLKTVLSIKNHLPSLNFTSPNPDLDLALRLV
HTATGPMSPDRPLVAGVSFGMGNTCHVLSLNRNAGSGAGKPTGTEDRLGAT
EAERKPDATNGPDPAQDTHRYPLILARSADAALRAQERLHHLHSHSGQRLDRT
AYSLATROVFERHVAHPGHDRELDLNGLENGLPAPQVLLGRTPTPEQGLAFLF
SGQSQOPGMRKRLHVPFPGFDALDEVCAELDTHLGRLLGPEAGPPLVWFARERT
AHSALLSETHYQALFALETALFRLLVQWGLKPDHLAGHSVGEIAAHAAGILDS
AAELVATGALMRSLPGGVMLSVOAPSEVAPEVLLGLREAHVGLAANVPDRAVVS
RGHVAATEQILDRGRKRYLRVSHAFSLPMEVLEFAEVAAGLTFRAPPTPLVSN
VTAAPEPCHRGEGADDAHTALLPALRRGRDEARSLTEAVARLHLGVPMWMTSVL
GGDVRVPLPTVAFQRESHMLPSGEAHPRPADDTESGTGRTGTEASPRPHDLHLVRS
AAVLGHSAERIDPDRAFDLGFDSLTALSELRLDTALGLRLPSSVLPHPSGAL
ARFLQGDTPPEKPGTNGTRATPEGPDDEPIAIVGMACRFPGVTSBDELWLLA
AGEBDAVDFTRGNNVTSATRRGGFLYDAGEFAAFPGISPREALVMDPOORLLE
TSWALERAGVSPGSLRGSDTAVYIGATQDYPRLHESDDSGGYLTGNTGTVASG
RIAYSLGLEGPATVDTACSSVALHVAQALRGECSLALAGGATVMPSPGMEVBF
SROGLSEDEGCKAFAATAGTGAEGVLLVERLSDBRLHVLAVRGSANVD
CASGLTANPGSQORVIRALADAGLPADVDVVEAHGTGTRIGDPIEAQALLATYG
CGRAGRPVILGSKNSIGHTQAAAGVAGVMKMLALGRGVVPTLHVDESPSHVDS
AGVELAVEAVPWSRGRRAGVSFGISGTHNAHVEEAPAEPEPEPPEPPEP
VVPVWSGRDAGALREQAALAAHVSGVADVGVMSLVATRSVEFHRAVMSSELDVA
AESLAGPAAGVVPVSGVAPAEGRVVFVPCGQSGOMVGMAAGLADACPVFAEVA
ECAALDPLTMSLVEVLGRGGAVALGRVDVQALWAVVSLARTWRYGVVEPAV
CHSGEIAAACVAGLSADGARVVVLRSAIRIAGGGVMSVSLPAGRVMTLDEEF
DGRVSAVANGPSTVSVQVQALDELLAGCEREGVRRVVFVSHASQMDQLRD
LLEALATVPTSANVFFSTVTDLTALDAGVFTNLTRETFVQFAGVLAQGM
GAFVCSPHVLPVGTITLDTFDADVALSLRRDEGLDRFLTSIAEAFVQGVVD
WSRAPEGASPTVDLPTVFORVYLLDKAQRERERLEDMRVHVEVRPVTPRPSR
LSGWAVALPARLADSLVCAIDALRGARVVPVVDREHQRQALVEALRNLGLD
DLGAGVSLALDEAPHGDPVPVMAASLALVQMAADAAPVFWFATGAVAAALP
GESPERQALLWGLRVVLEQPIWGLVLDLPQHLDEDRGLRVVVGGLADEDLQ
AVRASSVLAERLVTPGHRMSSQAGREWSPTVLTGTGTGALGARVWLAGKAE
HLVLSRRGADAAGAAALRDLSTGWRVTLAACDAADHALETLLDLSLRTDPAOLTA
VIHAAGALDDGMTTTLTPEOMNNAIRAKVTATVNLHELTRDLDSAFVLPSSISATLG
IPQANYAPGNSFLDAEWRRAQGLVATSIANGPWSGCTGMAHEGSGVERLQRHVL
AMEPAAATADHTLADSETAVADIDMSRFFLAYTALRAPRLIGEAPEARMLSEG
SGDGLPEFRAEPALAVLAGTAVEQLRLVQVQAQVVLGSHGAEPADPRAG
DLGFDLSLTVELRLNATATGLRPTAVFDYAPALAGHLRSLRIDDDGDHGLVDP
VKEHIDEPIAIVGMACRFPGIASPEDMDVLTAGEDVSVGLPNQRCMDLGRLLVDP
PDRTAGSYMREGAFLEHAGEFDAAFPGISPREALAMPQORLLLETWEALERAGITP
SKLAGSTGVFFGMSNQDQAQAGDVPSEGLYLTGSSISSVASGRVATYTGLEGPAK
TVDTACSSSLVHLAQGLRGECSLVGGTVMSPTVLTITFSRQSGVSDGRCK
AFAADAPFGAAGGVLLVLDARLGRHLRVLAVGSANODGASNAAPNGPS
QOVRIRALADAGLPADVDVEAHGTGTRIGDPIEAQALLATYCGRTSGRVPWLG
VKSNIHTQAAAGVAGVMKLVGRGVVPTLHVDPGSVPHVDSVAGLVAEAPV
SGRGRVRRAGVSFGISGTHNAHVEEAPAEPSVEEGPSVVPVWVSGRDAGALR
AQARLAAHVSTGACGVVDVGSVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGVV
PGVVGVAPEAGRRVVFVPCGQSGVMAGLDDACPVFAEVAECAVAAVLDRLTGWS
LVEVLRGGEAVLGRVDVQVQALWAVVSLARTWRYGVVEPAVVSCHGGEATAACVAG
GLSADGARVVLVLRSAIRIAGCGGMSVGLSABRVTRTMDTTCGRVSAAVNGPS
TVSGDAQALDELLAGCEREGVRRVVPVSHASQMDQLRDLLEALADYTPDSS
VPPSTVTADWLTALDAGVFTNLTRETFVQFAGVLAQGMCAFECSPPHVLVP
GITETLTDFDADVALSLRDEGLDRFLTSIAEAFVQGVVDVWTHAFEGEPFRPVD
LPTVFORVYLLHEEPDEADABFWSVVERGDAVPSDLSLLTDAALHTVLP
ALSSRRRRVREHRLQDMRVREKVPFPAALDEVLGGLWFLFVWGDGVAWRVVA
AVTARGGEVSVLELDPTRDRAVAAEAVAGRGVSVVSWDSDRRHSEHSVVPAGLA

```

RECTSASRLVFVTRNAVSDGGSVRDVAAMVGLVRSASQESPCRIALVDLEADSG
DLAEVLAGLPGILAAAGDAQFVVRGEGAVRGLVRESNEGAVSPREMDADGTVLVTGGT
GILAEVLARHLVVERGVRLHLLVSRGSEAGSGELRLDELAHGEIAVRACDVADRA
AVEAVGVADVAAHPUTAVIHTAGVLDGDIAGVIAISLTERLSAULVRPKVDAANLHEATKD
LDLAAFLVYSISGVLGSGQGNVYAGNVFIDLALAGHRAGLGPQASLAWGAWTQGG
MTATLSDMDQMMSYGTAPLSPEGLALFDAATGASATHLVAVRGVSQVPRACQGPV
SNMLRGVAGARTAVTGTETDHLARLRARPEERTQMTDVTVRTEAAALVIGHASS
QNVDAARFELVGLPSSLTVELNRSLSTVGLRLSATVVFDSPKNTDLDAAVYADLAA
QSALDGSASQAGFLVPAAPETDLSERMFIDLALDSCKVPEAQRMLSALCALPSPENTA
ELELDLPATLAEAGCTGTRLCVSTPTANGGVHESYARLAAAFRGRHVSALPLGVGFAT
GERLPATATVVRVIAESALRASDGNPFVLVGHESAGAPAYLAAALLENLTWIGRPEAV
VLDDTLSLRHEQNESIDVAGLRHFMVDVSPVMTNRSLSAMARWMGLNQLSEVRH
TTPVLLIIRATRETCIGTDTGIGEDHSGSPVDVRSVDADHFSWVRDDAPETARIVE
WLDLSGNA"
/complement (9772..15198)
/gene="pimS3"
CDS
/complement (9772..15198)
/gene="pimS3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/translation="MTTSETVVAALRASLKEAENLRQRNKLVAATEPIAIVAMAC
RYPGQSPEDLWELLTAGKDAIGAPPTDRGWDLDALDAGVDRRGNVAVSOBGGFLDG
VGDPADEFGIIPRESVMDPOORLLLETSEWAIERAGIDPTTLRSRTGTVFIGNGO
DYALLVRSADATGDTGIGIAASATSGRLSYTLGLEGPATVDTACSSSLVALHTAS
HALRAGCSLALVUGVNVNPSGSLMEFSRGGDLAAGRCFASTADTADGWSEGVG
LVRLSLDARRNGHEVLAVVRGSAVNDQASGFTAPSPSQORVIRQALAAAGLS
DWDVVAHGCTGTPDPTBAQALLATYQDQDPERPLLIGSVKSLNHTQAAAGAAGL
IKWILAMRHGVLPTLYADVPSSHVDWTSGEVELLTPQEWPKAPWRAGVSSFGIS
GTNAVHIVQASDEPEEVAEPGLAPAVVPLVMSAKTEEALAAQLARISSRGCTASPD
VGFSLASGRSEFEHRAVLAAQCPDGEELSEAGRAETSRPLAVLSFGQGAQRVHGRE
LVVRFPVFADELHVAHFDAGLRDMFGDAEGLDETGTQPALFAIEVALDFRLAESL
GYRPFVGGHSIGETAAAHVAVGLADETAVLRIAEELAAAGRKTRLPVSHAFPLME
VAAVLDFGVSIAGVAGPEAVVLAGDETVLRIAEELAAAGRKTRLPVSHAFPLME
PMLBDFRRVAEGLSYGAPOLPLVSNVTGELAVEELVCSPDYVVRHVRETRPADGVA
VAARLDFGVSIAGVAGPEAVVLAGDETVLRIAEELAAAGRKTRLPVSHAFPLME
GVDMSGIPAGTGARRVDLPTYPFHERYHPRPAVLSDGVAGAGLSAEHPLLGATLTL
ASGSEVLNTRLSLQTHPMLTDHAIGNVFFPGTGFLELAHAAADQACDRVEDLVLM
TPLVLAEBNGTQVLVGPADBAETRTISIHARPEDDPQPMWQAGGTLTSGERVDL
FGSTVMPQDVAVLEGGYEGTGPAFGQLRVSNSCGDGA YAEVALPDHVAEDAGA
GGLHPALDLAVLQSHRIAGVSADDOFLPFWRGVSLHAGGASVLVRVVKTGDDSMS
IAVADEGAPVLASAEAMWVAYRDCTGQAAARHGAERDELLSLEWVGPEVPPAGVR
CYSLNDELGVGASVNASLADLSGDEDFVVLVPSVGGGVDPAAAEHLTARTLGLSWS
GREGSGSLVFTVRNAVSDGGSVRDVAAMVGLVRSASQESPCRIALVDLEADSG
GDLAEVLARHLVVERGVRLHLLVSRGSEAGSGELRLDELAHGEIAVRACDVADRA
TGGELRELARHLVVERGVRLHLLVSRGSEAGSGELRLDELAHGEIAVRACDVADRA
AAVHALVAAPVAEHLPLTAIVHTAGVLDGDIAGVIAISLTERLSAULVRPKVDAANLHEATKD
DLDLAAFLVYSISGVLGAPGOANYAAGNVFLDALAHWMGOGLPQASLAWGAWTQDG
GMTATLSDMDQMMSYGTAPLSPEGLALFDAATGASATHLVAVRGVSQVPRACQGPV
PSVLKGLVGAARTAVNTGETAVSLADRLRARPEERTQMTDVTVRTEAAALVIGHAS
AKNVDAARFELVGLPSSLTVELNRSLSTVGLRLSATVVFDSPKNTDLDAAVYADLAA
EBAATGPALELRLDLSALAAGAPDARTAAVTRAAVTLAQILDLDRDAPAEDEGASVAE
REEASTDEIFAPIDNELGRHSDR"
/complement (15210..43733)
/gene="pimS2"
CDS
/complement (15210..43733)
/gene="pimS2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS2 protein"
/protein_id="CAC20921.1"
/db_xref="GI:12055072"
/translation="MDQEQKLRLKRASDLRRSRQVRQVGELEAAAREPIAIVGMSCR
FPGVRSPELNSGLAAGDGIITKSPADRQWELLAEADAGEFSGFLHDAAEFDAGFFG
ISPREALANDPOORLLLETSEWAFERAGIDPAVTRGTRGMFVGAMPQOEVYRGPDDVD
QGFALTGTTTISGRLLAYFFGAVGPAVTVDTACSSSLVALHLSLRQSGSLALA
AGVTVNSPTTFVFNROQGLSADGRCSFADSADGTGWSGEGVGLVLERLSEARNG
HEILAVIRGSAVNDQASGSLTAPNGPSQRRVTEQALVSARLSADVEDVVEAHGTGT
LGDPEVAQALLATYGGCRDDDRPLLIGSVKSLNHTQAAAGMAGVIMVLAAMRHGVLPL

```

```

RTLHVDAPSSHDVMTQGAVERLLTEHVWPOQEQPRRAGVSSFGLSGTNAHTIILEEAPA
APSPAGOEOTAPVAEPLAAGVPMWLVSGTRREALRAQAARLLEHLTRRQCLAPLDVA
YSLTAAGEDEHRAEFTAAAGPDGARAALTALADGTAPAGLVQDTRKSLKLAFLPAGQ
GSQRPNGRLEAARFPVPATALDEVLGHGFEGLEPLKOVLPFAEGTPEAALLDQTVG
AQPALFAI E VALVRLAESFGITPDLFAGHSIGETIAAHHVAGVPTLADAAALVLRGL
MOALPEGGASVLEATEEVEEVLPLEDRPKLSIAA NVGPAAVTAGADDDVTAVEHF
DALGRRTKRLRVSHPAFDLARRARRHLPRMGPDGVLSAMAQSCCRGHGHRRRR
TSPLYDHLHRAHVARFADLRLARRARRHLPRMGPDGVLSAMAQSCCRGHGHRRRR
LKHARRVPRQGPPEAETLTAALAGLHTRGVQWPEPYFGQTGARRADLPTTVAFORRYVT
KSLPATGDDVRAAGLGAHHPLLTAAVSVANSVSDGLLLTGLRSRRTHPLMADHAVRGT
VLLPETAFLLELAVRAGDEACGRVEELTLAALPLVLPREGGVQVQVWVSGSPASRRAV
SVHSRDPGPBELPMTQHAAGTLVAGEHHAGFDATWFPADATPDLLODFTARMADTGV
GYGFLPGLAAWRSGDDVYAEVALPDSGAASTESSAAGFLGHLPDLDLAAHAAAFVD
LGEDARGCLPFSQDVTLHASGSEARVRLTDPGDDAVALADVTTCGPAVASIGSLVL
LTDPEBGLIDAAHSLVRDALFELRMSSEARVRLTDPGDDAVALADVTTCGPAVASIGSLVL
VTA VPGDPAAPHA PDVLA PVAGTTEDDVVTSHVARTAHVLERLWLTDERVYASRL
VFTVRGALATDARQASPDVAAAVMGLVRAAQSHPGFEGLLDLPDGTTEPATTEPLLRA
LALSDEPQTAVRGTAVLTARLAPQAQAATEHWDPEGTVVLTGATGGLGRVLRHLV
AERGVRLHLLVSRGADADAGELMAELAAHGAQVTLAACDLADRAAVRLLSGVPAE
HPVTAVVHSAGSLDDGVIASLTPERIISAVLRPKADAVMHLHEATRDLDLAAFPVFSLS
SGTVGAAGQGNVAAANAFDLAALQRRAGLPLGLSLAWGPAWTAGMTGGLTDDLDLR
LARMGTPLATEEGIALFDDAATDAAVLLPTRIDLSVLRVQSEIPLMLRGLIRTPAR
RSVVSASETAVTLVQQLSRLAETDRREVLDLVAGOVAANVLGHAGTADIDPRRLREL
GFDSLTAVELNRSLAATGLRTAATVIFDHTVVDLAAHLDDLGLMSEAAADAPRT
Query Match 54.5%; Score 710; DB 1; Length 84985;
Best Local Similarity 73.7%; Pred. No. 5.8e-81;
Matches 956; Conservative 0; Mismatches 315; Indels 27; Gaps 3;
QY 1 GACCGATCGGCATCATCGGCATCGCATGCGTTCCTCCGGCGGAGTCCGTCCTCCGGAAG 60
DB 71258 GACCGGTGCTGATCGTCCGATCGATGCTGCGGTACCGGGCGGCGTCCGCGGAG 71199
QY 61 GACCTGTGGAGCTGGCGGCTCGGGCGGAGACCCCATCGGGCGGCTTCCCCACCGACCGC 120
DB 71198 GACCTGTGGCGCTGGTTCGGAGGGCGGTGAGCGGGTGTCCGACTTCCCCACCGACCGT 71139
QY 121 GGATGGCCACCGAAGACGCGTCAGCCCGCAGGACCCCAAGCGCCGCGGACGCTTCTATCCG 180
DB 71138 GGATGGAGCTGGAGAGCTCTACAGCCCCCGACCCCGAGGCGCTCGGCACCTCGTACAC 71079
QY 181 CAGGAGCGCGGTTCCTTTCAGCAGCGCGGCGCATTCGACGCGCGGCTTCTTCGGAATCAGT 240
DB 71078 CGCTCCGCTGATTCCTCCACAGCGCGCGAGTTCGACCCCGATTTCTTCGGGATGAGC 71019
QY 241 CCACTGAGGCACTCGCGCATCGATCGCAGCAGCGGCTGCTGCTGGAGACGCTCTGGGAG 300
DB 71018 CCGCGCGAGGCGCTGGCGCACCGACGCCCGCAGCAGCGGCTGCTGCTGGAGACGACCTGGAG 70959
QY 301 GCGTTCGAGCGCGCGGGAATCGATCCGCTGTCGGTACCGCGGCTCCCGTACCGGCGTCTTC 360
DB 70958 GCCATCGAGCGCAGCGGCATCGACCCGCGCTCGCTCGGGGCGAGCGGTACGGGCGTCTTC 70899
QY 361 GCGGCGCGCTCTCTCTTCGATACGCGCCCGCGTATGGAACACCGGCTCTCGGAGGGCGCC 420
DB 70898 GCGGCGGTGATGTACACCGACTACCGCGCACCTCTCTCTCGCGGACCACT----- 70850
QY 421 GCGGACGTGGAGGGCCACATCTCACCGGTACCGAGGCGGCGGCGGCGGCGGCGGATC 480
DB 70849 -----TCGAGGGCTACCGCAGCAACCGGCGGCGGCGGCGGCGGCTCGGCGGCGGTC 70797
QY 481 GCCTACAGCTTCGGGCTCGAAGGGCGGCGGCGATCACCGTGGACACCGGGGTGCTCGGCATCG 540
DB 70796 TCGTACACTTCGGTTTCGAGGGTCCGGGGTTCGAGGTGGACACCGGCATGCTCGTCGTC 70737
QY 541 CTCTGAGCGCTGCATCTCGGCTGCGAGTCTGCTGCGGTTCGGGTGAGTGACGCTCCCGCTG 600
DB 70736 CTGTGCGCCCTGCACCTGGGCGCGCAGTCTGCTGCGCTCGGCGAGTGTCTGCTCGCGT 70677
QY 601 GCCGCGGCGGTCTCGGTCTATCTCCACCTCGGCTCGGCTGTTTCATCGAGTCTTCCCGGAGCGC 660
DB 70676 GCGGCGGCGGTGACCGGTGATCTCCACCGAGCGAGTTCGTCGAGTTCCTCGCGGCAACGC 70617

```


Best Local Similarity 73.7%; Pred. No. 8.5e-81;		Matches 956; Conservative 0; Mismatches 315; Indels 27; Gaps 3;	
QY	1	GACCGATCCCATCATCGCATGGCATGCGTTTCCCGGCGGAGTCGCGTCCCGAAG	60
Db	9895	GACCGGTGTGATGTCGCGCATAGCTGCCGTACCCGGCGCGTCCGCTCGCCGAG	9954
QY	61	GACCTGTGGAGCTGGCGCTCGGCGGAGAGCCATCGGGCGTTTCCCGACCGACCGC	120
Db	9955	GACCTGTGGCGCTTGTGTCGGAGGGCGTGACCGCGTGTCCGACTTCCCGCCGACCGT	10014
QY	121	GGATGGCCACGAAACAGCTGACGGCCAGGACCCACGACGCGCGGACGTTCTATCCG	180
Db	10015	GGATGGGACGTGGAGAGCTTCTACGCCCGACCCCGAGGCGCTCGGCACCTCGTACCC	10074
QY	181	CAGGGAGCGGCTTCTTACAGCAGCGGCGCATCTCGACGCGGCTTCTTCGGAACAGT	240
Db	10075	CGCTCCGCTGGATTCCTTCCACAGAGCGCGGAGTTTCAGCCCGATTCTTCGGGATGAGC	10134
QY	241	CCACGTGAGCACTGGCGATCGATCCGACAGCGGCTGCTGTGGAGAGCTCTCTGGGAG	300
Db	10135	CCGCGGAGCGCTGGCGACCGACCGCCAGCAGCGGCTGCTGTGGAGAGCACTGGGAG	10194
QY	301	GGCTTGGAGCGCGGGAATCGATTCGCTGTCGCTGACGCGGTCCTCGTACGGCGCTTTC	360
Db	10195	GCCATCGAGCGCACGGGCATCGACCGCGGCTGCTGCGGGCGACGCGTACGGCGCTTTC	10254
QY	361	CGCGGCGCCCTTCTTCTGACTACGCGCCGCGTATGACACCGCGTCTCGGAGGCGCC	420
Db	10255	CGCGGCGTGTATGACCGACTACGCGGACCTCTCTGTCGCGGACCGAGT-----	10303
QY	421	CGCGAGCTGGAGGCGCACATCTTACCGGTACACGGGACGCGTCTGTCGGGCCGTATC	480
Db	10304	-----TCGAGGCTTACCGCAGCAACGGCAGCGCGGCGAGCATCGCCTCGGCGCGGTC	10356
QY	481	GCTTACAGCTTCGGGCTGGAGGCGCGCGATCACCGTGGACACGGGGTCTCGGCATCG	540
Db	10357	TCGTACACCTTCGGTTTCGAGGGTCGCGGCGTACCGTGGACACGGCATGCTCGTCTC	10416
QY	541	CTCGTACGCTGCATCTGGGTGCCAGTCTGCTCGGTCGGGTGAGTGCAGTTCGCGCTG	600
Db	10417	CTGGTCGCTGCATCTGGGCGCGCAGTCTGCTCGCTCGGCGGAGTCTCGTCTCGGCTG	10476
QY	601	CGCGGCGGCTCTCGGTCATGTCACCGCTCGGCATGTTTCATCGAGTCTTCCGCGCAGCGC	660
Db	10477	CGCGGCGGTGACGCTGATGTCACACCGACAGCTTCTGTCAGTCTTCCGCGCACCGC	10536
QY	661	GGCTCTCGGTGACCGCGAGGTGCAAGGCTACTCGGTCGACGCGACCGCGCTCGG	720
Db	10537	GGACTCTCGCGGACCGCGCGCTGCAAGGCGTTCGCGCATGCGCGCGACGCGCGTCTCG	10596
QY	721	GGCGAGGCGTGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCTCGGTGCGGCTGGGCTAT	780
Db	10597	GGCGAGGCGTGGGATGCTGCTGCTGAGAGCGTCTGTCGGAGCGGCTGTCGGAGCGCGCC	10656
QY	781	CGGCTCTCGCGGTGCTGACGCGGCTGCGTCAACAGGACGCGTTCGATCGAGTGGGCTG	840
Db	10657	CGGCTCTCGCGGTGCTGCGGCGCATGCGGTGAACAGGACGCGTTCGATCGTCTGCTG	10716
QY	841	ACGCGCGGAAACGCTCCGCTCAGGACGCGGTGATCCGCGACGCGTTCGCGAACCGCGGG	900
Db	10717	ACGCGCGGAAACGCGCGCGCCCGACGACGCGGTGATCCGCGACGCGTTCGCGAGTGGGG	10776
QY	901	TTGTCGCTGGCATGCTGATGTTGGTGGAGCGGTCGCGGACCGGCGACGCGTTCGCTGAT	960
Db	10777	CTGTCGCGCGGATGTCGACCGCGTGGAGGCGGACGCGTACGCGGTACGAGCTGGCGCAT	10836
QY	961	CGCATCGAGGCAAGCGGCTTGTCTCGCACGATACGG---CGACGCGGCGCGGTGACAGCGCG	1017
Db	10837	CCGATCGAGCGCCAGCGCTGCTCGCCAGATGATGGCCAGGACGCGCTGAGGACCGCGCG	10896
QY	1018	CTGTGCTGGGTCTCTGAAGTCCAAATCGGCGCACCATGCTGCGCGCGGTGTGGT	1077

Db	10897	TTGCTGCTGGGTGGTCAAAATCCAAATCCAAATCGGTCTATGCGACGCGCGCTTCTGGGTGTGGCG	10956
QY	1078	GGGGTCATCAAGATGATGCGTTCGCGGAGGGGGTGTTCGCGCGGACGTTGTCATGTG	1137
Db	10957	GGTGTATCAAGATGATGCTGGCGATGCGGACGCGTGTCTGCTCGGACGCTGCTATGTG	11016
QY	1138	GATAAGCCGTCCCGCAGGTGACATGCTCCGCGGGCGCGTTCGCGTCTGCTGACGAGGCG	1197
Db	11017	GATGAACCGTCTGTCGATGTCGACTGAGTGCCTGCGGTGCGTCTGAGCTGCTGACCTCCGAG	11076
QY	1198	GTGCGGTGCGCGGAGCGCGGCGGCGTTCGCGGCGCGGCGGAGTGTCTGTTCCGG	1257
Db	11077	GCGGAGTGGCGCGCGCGGCGG-----GGCGCGCGCGCGCGCGCTCTCTCTTCGGC	11130
QY	1258	ATCGGGCGCACGAATGCGCATGTGATTTTGGAGGAGGC	1295
Db	11131	GTCAGCGGACGAACCGCATGTGATCCTGGAGACGC	11168

RESULT 10
AX067996
LOCUS AX067996 20394 bp DNA linear PAT 19-JAN-2001
DEFINITION Sequence 3 from Patent WO0077222.
ACCESSION AX067996
VERSION AX067996.1 GI:12329806
KEYWORDS Streptomyces natalensis.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
AUTHORS Martin, J.F., Aparicio, J.F. and Colina, A.J.
TITLE Genes encoding enzymes in the biosynthesis of pimarin and the application thereof
JOURNAL Patent: WO 0077222-A 3 21-DEC-2000;
DSM N.V. (NL)
FEATURES
source
1. .20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
1. .20394
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC24642.1"
/db_xref="GI:12329807"
/translation="MSNEEKUREYLKRAIDLHETRQQLDETEAKOREPLATVSMACR
FPGVRSPEELWELLRDGVDVAVSPFNRGMDLDALYHSDPAHQSTVAREGFLHDA
GEFDGFGISPRREALMDPQRLILLETAWERAGIDPESLAGRTGVFVGTGHG
YDAERRRDEVGHHLTGNHISIASGRISYVIGLEGPALTVDVTACSSSLVALHLMH
ALVRLDECAMALVGGAATVMSVQMFVFSRQSLAPDRCKPAAADCTGWSGVCLL
LVERLSDAVRNGYFVLAVLKGSAVNDQASNGLTAPNGSQORVTRQALTGAGLAASD
IDAVAHGTTGTLDFVEAHALLATYQOORADRPCGLGSMKSNIGHTQAAAGVGM
KMLVAMRHGLPTGLDDEPTCHVDVSEGNARLAEFPWPSAGRPRAAASVFGISG
TNAHVLQAPAEHAEPAPAPALPWLISARTEAGRLAQRDLGRHLRADLE
PAAVAHALDRTTLMHRAVAVAGREFFLGLDALAAGRTANGVSVGVAKAAAFLL
FAGGOSORPGMREHLHAPVFAFADVACAEELDPLDRPLDIIVAEBSGAALLD
QTAQALPALETALFRLVSGVAFVAGHSIGELTAHVSGVLTILHDSARLVA
RGTLQALPAGAMVAVQADEIRERLAGHEDHVALAAGPDSTVLSGDSQAVTEI
AAHWAQGRTRKRLRSHVAFPHMDLDEFRVARGLTTHAPRIIPVSTVGTALAT
DELRSPDYVRQVRETFVCAAVTLEAGVTVEIGTGVLTVMQVDCLTLEEP
VLVPLLRTRPETVALTEGVATFVHGVVDVRSAPFAGPAGTSRADLPYAFORVWL
DPADHGEAAAEAGFWAVERDLOELSLAVLADGSEADSGFLPPLTSSWR
QRRQAADRFSYRTHAPRTASGPTATGHLVVLPEGGTDDPTWALLDALNQQGL
HTDVRELPADHPDPAWGRHPVDVLCCLLADDERPTRCPYRRRGLAATTAARPEGA
GTQPLWCVTGAVADHEALKSPLOQTWGLRVALESQSMGGLITDPNLDGR
AVSALLSTLAGEEQVAVRPGVFAARLERITPGDGTGRWSTHCTVLTGTTGALG
HLAWLADAGEHLVLCGRGQAPAPALAAATDGRVKVTLAACDAADRDLAAVL
ADIPHLPLTGTVHAAVLDGVLDTPEFETVLRPKAAONLHELTDLDLHP
VTPSSIVGLNAGQANYAANAVALDAEHLAQGLPATSVSWNGPQGAWHSDAA
DRMSRDLGLPMAAARRRRPALAQMTQVTVADLWSAPALPAPVPSPLIGDLE
ARRALGAPGRRERSRDRIGALPPAEQKAFVTMREEARVILHPSPTVDAQR
AFREGFDSLMAVDLRNLSAATGLRLPATLLFDHPTPLAAAAACLRSEVLGAAGPATV


```

/db_xref="GI:21999186"
/translation="MGRQYQVAVGSLRREERGLRFLTSAAEAFFVQGVDTWPALEFDDT
GARAIIDPTYPFQHQNYWYIDPARRKGDVSSFGMAEAGHPLLDAGTELPEGGEHLTYTA
RLTADSHVPLLEGLTLLPCAAFDVILVWAGGELGELVEELFLTSPLLLSDSAA
VQLRVLVGATDABRGRTITTHSPDPAHPTGTRTPAAPPEAGSEAEVRDAADWTSHAQ
ATVAPAPASVVGADGNPGLPAASWSSAATESDTRFADFFYASFAARGYTGELFLQ
GVRSGRQSDVVAEALDQALPPDGERFGLHPALLDAAAFQTNRLGSPFFPDQARVP
YTRGRLIYAPGARLRVSVAVGDDAVECADEQRLVCEIDALVLSVSPQAQLRP
AGQDATALMRLHVPSPAGRATSTAAFRWIVVGPEDDGLGLAEVSGTGTER
HRLDALTRALTPTDPPGSPAADKTLPIAVPSPSPMDAVSVRHAHQALVESQVQ
WLSADEACGSGGAGAGRLVLTSGAVSTGDADPLDPVPASAVWGLVRSQSEOP
DRFVLVDLEGADGALSAIASGEPQALAREGKRVHVPALASDRMRLLPFGAG
RAWHLVKATGTLSDALVPAQATVPLAPQVRIAVRAAGLNFRDILLALGMYPEEG
VMAEGAGVITEVPGVAGFVGDVLMWTDGFPYAVADHRMVAEMPRDWSYAEAA
SVPAFLSYGYGLKHLAGRAGQSVLHAAAGVGMAAQVQLARYFAEYVGTAGPAKV
DVLAOGLDDRHIASSRTLDFAARFLDATTGGGVVVLNSLAGDFVDAISRLLPGRGR
FVELGKADYDRAQVAADPCTGYRAFELMEAGPELIGMLTELELFPSSGALRLPVS
TPYDIRPADAFITLSQAHVGLVLTMPADPPRGPACSCRPFRHTRRRGTNSA
ARPGAPARCRITLVRCGHAGCGGPGAPDHGRIHPRGTARREARTRARPQAGARA
HNRLLGARAGRPTPTVLLQSGRIHLAHRRTPOQAHRRHNEASRHTGLRPQDAG
GTRRPPGTGAEGRSRSTRGRPCRSGTSDRDPORTRHRRQRPRPRRGTPSGTARH
ARPAGGRRPPGRLRSRRAGR"
36783..37874
/notes="orf5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAM81588.1"
/db_xref="GI:21999187"
/translation="MNDEWLLYRGVEPHDGIHALPPPPWRDFFDGGPAEAPADAAG
EGGAAGAGGGGPDADGNVRLRGAHRAQAAHREPEELEEANALYLRPLPLVT
GIPGTGSTLAHVAHELKLRVLPVVSRTVLOEGLRYDAIARLQDVQIAAGGVA
GAAGAGSAGSGSGSIGRYIRIGPLGTALLPTARPRVLLIDELDKSIDLNDLLN
VLEEGFTPELERVADTEPEVRLTDDCAKVTRDGRVRCRAFPIVLTNSGERDFF
AALLRCIQLKQGPQEKRLATWRAHLEBEAARLIGADLIREFLRSRTELVAADQLL
NAILYTHAAPFTREDLADLLIQLDRPR"
BASE COUNT 5228 a 14409 c 16025 g 6119 t 1 others
ORIGIN
Query Match 55.5%; Score 722.2; DB 1; Length 41782;
Best Local Similarity 73.3%; Pred. No. 2e-82;
Matches 954; Conservative 0; Mismatches 338; Indels 9; Gaps 2;
QY 1 GACCCGATGCCATCATCGCATGGCATGCCGTTTCCCGGGGGAGTCGCGTCCCGAAG 60
DB 21885 GAGCCGATTGCGATCGTGGGATGGGTGCCGCTCCCGGTGGGTGTCGTCCCGGAG 21944
QY 61 GACCTTGGGAGCTGCGCGCTCGGGCGAGAGCCATCGGGCGTPTCCCAACCGACCGC 120
DB 21945 GGGTTGTGGATCTGGTGGCGTGGGGGTGGACGCGGTCTCTCCGTTCCCTACCGATCG 22004
QY 121 GGATGCCACGAGAGCTACGCCACGCGGCACTTCAGACGCGGCTTCTTCTATCG 180
DB 22005 GGTGTGGATGTGGCGGCTCTGTTTATCCGAGCGGGGTGTCGTCGGCGCTACGTATG 22064
QY 181 CAGGAGGCGGGTTCCTTCACAGCGCGGCACTTCAGACGCGGCTTCTTGGATCACT 240
DB 22065 CGTGGGTTGGTTCCTGATGGGCGGGGAGTTCGACGCGGGTTTTTCGGTATTCG 22124
QY 241 CCACTGAGGCACTGCGCATGATCCGACAGCGGCTCTGCTGGAGACGTCTTGGGAG 300
DB 22125 CCGCGTGGCGTGGCGATGATCCGACAGCGGTTGTTGCTGGAGACGTCTGTTGGAG 22184
QY 301 GCGTTCAGCGGGCGGATCGATCCGCTGTCGTACGCGGTCGCGTACGCGGCTCTTC 360
DB 22185 GCGTTCAGCGGGCGGATCGATCCGCTGTCGTACGCGGTCGCGTACGCGGCTCTTAC 22244
QY 361 GCGGGCGGCTCTCTCTTCTGACTACGGCCCGCGTATGACACCGCGTCTGTCGGAGGCGC 420
DB 22245 GCGGAGTGTATGACACGACTACGCGACACCGTGG---CGGGTCCGCGCTCCGAC 22301
QY 421 GCGGAGCTGAGGCGCACATCTCACCGGTACAACGCGGAGCGTCTCTGTCGGCGCGTATC 480

```

CDS

```

DB 22302 GACGAGACGGCGCGGTTCTCTCGGCATCGGCACGTCCGGCAGTGTCTCGGTCTGATC 22361
QY 481 GCCTACAGCTTCGCGGTGGAAGCGCGCGGATCACCGTGGACACGCGGGTCTCGGCATCG 540
DB 22362 TCGTATGCTGGGTCTGGAGGGGCGCGGTGACCGTGGACACGCGGTGCTCGTCTG 22421
QY 541 CTCGTGACGCTCCATCTGCGCTGCGAGTGCCTGCGGTGCGGTGAGTGACCGTCTCGCTG 600
DB 22422 CTGGTTCGCTTCATCTGCGCGTGGGGCTTTTGGCAGTGGTGAAGTGTATCTCGCTCT 22481
QY 601 GCGGGCGGCTCTCGGTATGTCACCCCTCGGCATGTTTCATCGAGTGTCTCCCGGACGCG 660
DB 22482 GCGGGTGGGTACCGGTCTATGCGCGAGCGGGGTGTTTTCTGAGAGTCTCTCGCGCACG 22541
QY 661 GGGCTGTCTGTCGACGCGAGGTGCAAGGCGTACTCTCGGTGCGAGCGACGCGCACCGGCTG 720
DB 22542 GGTCTGGCGCGCATGGGCGGTGCAAGCGGTCTCGGACTCGCGCGACGCGGCTG 22601
QY 721 GCGGAGGCGTCCGGATCTGTTGTGTGAGCGGTTGTCCGATGCGGTGCGGTGCGGCGCAT 780
DB 22602 GCCGAGGCGTGGGTGTGCTGCGGTGGAGCGGTTGTCCGATGCGGTGCGTCAAGCAT 22661
QY 781 CCGGTGCTGCGGTGCTGACGCGGCGAGTGGGTCAACAGGACGGTGCGTCTGATATGGCTG 840
DB 22662 CCGGTGCTGCGGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 22721
QY 841 ACGGCGCGAAGCGGTCCGCTCAGGAGCGGGTGATCCGCGAGCGGTGCGCGAACGCGGG 900
DB 22722 ACGGCGCGAAGCGGTCCGCTCAGGAGCGGGTGATCCGCGAGCGGTGCGCGAACGCGGG 22781
QY 901 TTGTCCTGGCGGATGTGATGTTGTGTGAGGGGCGACGCGGACGCGCACGCGTCTGGTGT 960
DB 22782 TTGGGTGTGGGTGATGTGATGTGTGTGAGGGCGCATGGGACGCGGCGCGTCTGGGTG 22841
QY 961 CCGATCGAGGCGACAGCGGTGCTGCTCCACGCTACGCGGCGAGCGGCGGTGACAGGCGCGT 1020
DB 22842 CCGATCGAAGCGCGAGCGGTGTTGTGGCGAGCTATGGGCGAGCGCGCGGCGCGCTTTG 22901
QY 1021 TGGCTGGGCTCTCTCAAGTCCAAACATCGGCGACACATGGCTGCGCGGTGCGGTGGG 1080
DB 22902 TGGCTGGTTCGCTCAAGTCCAAATGTTGGTTCATACGAGGCTGCTGCGGTGCGGTGG 22961
QY 1081 GTCTCAAGATGTTGATGCGGTTCGCGGAGGGGGTGTTCGCGCGGACGTTGTCATGTGGAT 1140
DB 22962 GTGATCAAGATGTTGATGCGGTTCGCGGTGAGGTCTCTCCGAAACGCTGTCATGTGGAT 23021
QY 1141 AAGCGCTCCCGCAGGTGATGCTGCTCCGCGGCGGCGGTGCGGCTGCTGACGAGGCGGTG 1200
DB 23022 GAGCGCTCCCGCAGGTGATGCTGCTCCGCGGCGGCGGTGCTCTTCTGACGAGCAGGAG 23081
QY 1201 CCGTGGCGGGGAGCGCGAGGGCGGTTCGCGCGGGGCGGAGTGTCTGTTCCGGATC 1260
DB 23082 CCGTGGCGG-----GAGTGGGGGCTCCGCGAGGGCGGCGGTGCTCTCTTGGGCTC 23135
QY 1261 GCGCGCACGAATCGCATGTGATTTTGGAGGAGGCGCGCGG 1301
DB 23136 AGTGGAGCAATCGCATGTGTTGTGGAGGAGGCGCGCGG 23176

```

RESULT 9
SNAL32222
LOCUS Streptomyces natalensis pims1 gene.
DEFINITION Streptomyces natalensis pims1 gene.
ACCESSION AJ132222
VERSION AJ132222.1 GI:4678702
KEYWORDS pimaricin biosynthesis; pims1 gene; polyketide synthase.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
AUTHORS Aparicio,J.F., Collina,A.J., Ceballos,E. and Martin,J.F.
TITLE The biosynthetic gene cluster for the 26-membered ring polyene

LTCASGALGCLVARHLVAEHGVRSLLLLRRGAEPAGAVELEBAELAAWGAEVVRAACD
 VADREAVGMLNGLSQSLASAVHTAGVLDGIVASLTPTERRREVFRPKVDVAVLNHLE
 CTRDMGLAFAFVSSVAGVAGSAYAAANSFLDAFSAHRRRREGGLPAISLAWGYWE
 QSGMTDGLVDEARMAWSVLPPLPEGLRLFDAALASDEAVLAPVLDVTCALRAGL
 EAPVLVRLPAAARRTARAAAAPSASLAELEEMPEAREKTVLDIRVAEVRAGL
 GHASDRTPPEHAFQDLGDSLTAVELRNRLNKATGLRLPATILVDFDHPPTALLARHFV
 MAAGAABEPALVDLADLDREVEQLVTTILASEANDRILSLQGLLITAGEBFGKAPA
 QEGPFGGSDLESATDDEVFDLLGKEFGIS"
 21273..31847
 /gene="pteA2"
 21273..31847
 /gene="pteA2"
 /codon_start=1
 /transl_table=11
 /product="modular polyketide synthase"
 /protein_id="BAB69304.1"
 /db_xref="GI:15824140"
 /translation="MSCRYPGVDSPEDILWELVAGGRDAVSEFPADRGWDLDALYDPP
 PSRGTSYAREGAFLRDARFDAGLFGISPREALAMDPOORLLLETSEWELFERAGIDA
 ASLAGSRTGVFAVWNHFEFLAALQNSPQLEGLYGTGTSVSGRVAFTFGLGPAV
 TVDTACSSSLVHLHVAQSLRNGECSLAVAGVTAMVGPATVFGSRQRLGAPDRCK
 PPAAGADGTTWGEAGLLVERLSDRNRGHEVLA VVGSAVNQDASNGLTAPNGPS
 QQRVIRQALAAQLTTAQVDAVEAHGTGRLGDPTEAQAALLATYCGRPDDKPLWLS
 IKSNGTQAAAGVAGIIKVMAMRHGELPRILHVDAPSVDVSDVWDTAGAVRLLETPTW
 SASDQRRRAGISFSGTNAHTIIEAPSAEAFADQDTGAPEPEEGAEAPVASEA
 ARPAGGPVFWISGRTEALRAQARLLSRDABVPLDVAFLATTRLTALEHRAVL
 QETPAELAEGLRVASGSPARGVGLSGTGVGLFSGOGSORLGMRLGYASYP
 AFVAAVDEYCAQLDAVVDVDSBELHRTGAAPALPAVEVALFRLLESWIGIRADYVAGH
 SVGELAAHVAVLSDAALKVBARLMOALPAGGAMVAQATEDEVLPHLTAVAG
 IAAINGPRSVVSGREDAVTAIAEPRQGRKSRLLKVSHPASHPLDILSEFAEV
 RLUTHFRPOIPVNSVTLRLABPYTPQYVWRHVREAVRFADGVRTLNDLGVTFVFI
 PGVLASAEGLDLDVVTPALRADRPVVALVAASSELHAHGVSPDWHAFPPGAR
 VDLPTAFVORYERWIDLAPPAAGVRAAGLADHPLIGAAVLAGGDELLTGLRLSL
 RTHPLMSDRAVLGTLALPCTAFVELAVRAADEAGCALLDLTLEAPLVVPEGGVAVQ
 VVWVAVDGTGRPLTVHSRPEDSDLPVWRHATGFTVTEGSEAGAAADRLTANPPT
 56.7%; Score 738.6; DB 1; Length 78210;
 Query Match
 Best Local Similarity 74.7%; Pred. No. 1.5e-84;
 Matches 975; Conservative 0; Mismatches 309; Indels 21; Gaps 3;
 QY 1 GACCGATCGCATATCGGCATGCGATCGCGTTTCCCGGGCGGAGTCGGTCCCGGAG 60
 DB 15790 GAGCCCATCGCATATCGGCATGCGCTCCCGTACCCCGGTGGCGTGCAGTCAACCCGAA 15849
 QY 61 GACCTGTGGAGCTGCGCGCTCGGGCGGAGCGCCATCGGCGCGTTCCCGCCAGCGCC 120
 DB 15850 GACCTGTGGAGCTGCTGTTCTCCGCGCGGAGCGCGTGTGGGCTTCCCGAGGCGCC 15909
 QY 121 GGATGCCCCACGGAACAGCGTCAAGCCAGGACCCACAGCGCGGCGCATGTTCTATCCG 180
 DB 15910 GCGTGGGAGCTGAGCAGCTTACGACCCGGAATCCCGACGAGTGGGGACACAGTACACC 15969
 QY 181 CAGGGAGGGGGTTCCTTACAGCGGGCGACTTCGACCGCGGCTTCTTCGGAATCAGT 240
 DB 15970 CGAGAGGGCGGGTTCCTGCACACGCGCGGCGACTTCGACCGCGGAGTCTTCGCGGATCTCG 16029
 QY 241 CCACGTGAGCGACTGCGCATGATCGCGAGCGGCTGCTGCTGAGACGTCCTGGGAG 300
 DB 16030 CCGGTGAGCGCTGCGCATGACCCGCGAGCGGCTTCTGCTGAGACGTCGTGGGAG 16089
 QY 301 GCGTTCGAGCGGCGGGAATCGATCCGCTGTGCGTACGCGGGTCCCGTACGCGGCGTCTTC 360
 DB 16090 GCGTTCGAGCGGCGGATCGACCGCGGCGACGGTGGGGCGAGCAGGACCGGTGCTCTTC 16149
 QY 361 GCGGGGCGCTTCTTTCAGTACGCGCGCGGTATGGACACCGGTGCTGTCGAGGGCGCC 420
 DB 16150 GCGGCGGTGATGATACGACTACGCGCGGCGGGTGCACACCTCCCGCCCGAGGGC----- 16204
 QY 421 GCGGAGCTGCGAGGGCGACATCCTCAACCGGTACACGGGCGAGCGTCTGTGCGGCGGTATC 480
 DB 16205 -----TGGAGGGTACTCTGTAACGCGAGTGGCGGCGAGCATGCGCTTCGCGCGGGTTC 16257
 QY 481 GCCTACAGCTTCGGGCTGGAAGGGCGGCGGATCAACGCTGGACACAGGGGTGCTTCGGCATCG 540

DB 16258 TCGTATACAGTTTCGGGCTCGAGGGCGCGGTCACCGTGCACACGCGGTCTCGTCTCG 16317
 QY 541 CTCGCTGAGCGTCACTTGGCGTGCAGTGCCTCGCTCGGTGAGTGACACGCTCGCGCTG 600
 DB 16318 CTCGTAGCCCTCCACTCGCTCGGAGGCGCTCGGAGCGGCGAGTGCTCGATGGCCCTC 16377
 QY 601 GCGGCGGCGCTTCGCTCATGCTCCACCTCGGCGATGTTTCATCGAGTCTTCCCGGCGAGCGC 660
 DB 16378 GTCGCGGCGCTCAAGTGTGATGCGAGCGCGCTCGGTTCGTGAGTTCAGCCGTGAGCGC 16437
 QY 661 GGGCTGTTCGCTGCAACGCGAGTGCAGGCTACTCGCTGTCAGCGACGCGACCGGCTGG 720
 DB 16438 GGCCTGTTCGCGCGACGCGCGTGCAGGCGTTCGCGCGCGCGGCGCGACGCTACGCGCTCG 16497
 QY 721 GCGGAGGCGCTCGGATGCTGTTGTTGAGCGGTTGTTCGATGCGGTGCGGCTGGGGCAT 780
 DB 16498 GCGGAGGCGCTGGGATGCTGTTGTTGAGCGGTTGTTCGACGCTCGTGCAGAACGCGCAC 16557
 QY 781 CCGGTGCTGGCGGTGCTACCGCGCGAGTGCAGTCAACACAGACGCTGCGTCGAATGGCGCTG 840
 DB 16558 AAGTGTCTGGCGGTGCTCGCGGTACGCGGTGAACAGGACGCTGCGTTCGAACGCTTTG 16617
 QY 841 ACGGCGCGAAACGCTTCGCTCAGAGCGGCTGATCCGCGAGCGCTTGGCGAACCGCGGG 900
 DB 16618 ACGGCTCGAAACGCGGCCCTCGCAGCAGCGGCTGATCCGCGAGCGCTGGCGAACCGCGGC 16677
 QY 901 TTGTCGCTGGCGGATGCTGATGCTGAGGCGGCGAGCGGCGGCGACGCGCTGGGTGAT 960
 DB 16678 CTGTCTTCGATGAGGTGATGCTGCGGTGAGGCGCGATCGGTTACGAGGCTGGGTGAT 16737
 QY 961 CGGATCGAGGCGACAGCGCTGCTGCCACGCTACCGGCG---AGCGGCGCGTTCACAGGCGG 1017
 DB 16738 CGGATCGAGGCGGAGCGCTGATCCGCGAGTACGCGGCGAGGCGGTTCGAGAGACCGCGCG 16797
 QY 1018 CTGTGCTGGGCTCTCTGAAGTCCAAATCGGCGCACCATGGCTGCCCGCGGCTGTGGGT 1077
 DB 16798 TTGTGCTGGGTTCTCTGAAGTCCAAATCGGCGCATTCGAGCGGCTGCGGCTGTCGT 16857
 QY 1078 GGGGTCACTAAGATGCTGATGCTGCGGAGGCGGCTGTCGCGGAGCGTTCGCGGAGCGTTCGATG 1137
 DB 16858 GGTGTGATCAAGATGCTGATGCGGATGAGCAGCAGCGCTGCTGCCCGACACCTTCGATG 16917
 QY 1138 GATAAGCGCTCGCGGAGTGCAGTTCGCGGCGGCGGCTGCGGCTGCTGACGAGGCG 1197
 DB 16918 GACGAGCGGACCCCGGACGCTGAGTTCGCGGCGGCGGCTGCTGACCGAGGCG 16977
 QY 1198 GTGCGCTGCGCGGCGGACCGCGGCGGCTGCGGCGGCGGAGTGTGCTGCTGTCGCG 1257
 DB 16978 GTGAGTGGCTGCGAGCGAGCG-----GGCGCGCGCGGCGGCTGCTGCTTCGT 17031
 QY 1258 ATCGGCGGCGCAATGCGCATGTGATTTTGGAGGAGGCGCGCGG 1302
 DB 17032 GTGAGCGGCGCAACGCGCACGTCACTGTCGAGGAGGCGCGGACG 17076

RESULT 8

AY118081

LOCUS

DEFINITION

Streptomyces sp. GER1155

41782 bp DNA linear

BCT 27-JUL-2002

and unknown gene.

ACCESSION

AY118081

VERSION

AY118081.1

GI:21999182

KEYWORDS

ORGANISM

Streptomyces sp. GER1155.

Streptomyces sp. GER1155.

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 41782)

Lim,S.K., Nam,D.H., Sohng,J.K. and Yoo,J.C.

Cloning of polyketide type 1 synthase gene of Streptomyces sp.

GER1155 producing a macrolide antibiotic, GER1155

Unpublished

JOURNAL

REFERENCE 2 (bases 1 to 78210)
 AUTHORS Ikeda, H.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel.: +81-3-5791-6242, Fax: +81-3-3444-6197)

FEATURES
 source
 1..78210
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 /note="Actinobacteridae"
 complement (1..1254)
 /codon_start=1
 /transl_table=1
 /product="putative lipoprotein"
 /protein_id="BAB69301.1"
 /db_xref="GI:15824137"
 /translation="MGEALITLSTRSLAATAGVLGGVLAATVLTALGGTSNAATSGNGA GNHPTSOARINQAGAEASDARIEITRCQGTYSVCTNHPVSVTVTSGLTKVMTTAVA TCABIPGLSTGCTGSKNGLERATKQIAABAGDVKGRSTGTSATITVSPANDFI GHLSPEGSTGVTNLRHGIQKTVFKIGRSQISTVDKTKMTVVRDKTKITPIS SGSPHPTNGQMTVSEKFRQIHMGASVGLTEKQKPSYDIKAVPHAMRLTDSGTFI HGNYGTDVSCNANTSHGCVGLKDVGRGDHGRQAPAAFFEDHSIVGDVIVKNSSEKT MLAPDNGLSDNWNSSEWAGSPAT"
 complement (1256..2290)
 /codon_start=1
 /transl_table=1
 /product="putative transposase"
 /protein_id="BAB69302.1"
 /db_xref="GI:15824138"
 /translation="MPVFEIHSQDVSRVGRFASAPTPGELEQFPRLDAXALESARSKR APANRLWAGVQMGCVGFTEDLSVVEVVVFAAEQGVDPGEAAFGARRQNR YEHWEIADAYVRAEPAEAEFLATRVWMEGPRALFDRAVGLVNRVLLPG ITTLRLVAEVAQNAALYRTLDEAVPEKLQSMRDLKVPGRVSELERLRSPM GYSGADSLAFSEIRMEAVMTAGPGSTDIKELRVSVQSVQRWGLACDRGPRG AVQPVSVELSSTLPAALDEERAKGAPSWTIRYLHDDPPIFGQRYAYKWSQGLHSP GRSAQGRDQ"
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=1
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translation="MVSEELNATGNRDCSASADSADGIAVIGLCRLPGAADPAAPWR LLSREGADATDVPDWDGCAVADASAPDRGTGIRGGFLDSIGDFADGFFGLSPKEAA ANDPQORVLELAWALEDAHVRAATLSTRGTGVFGAMDDYATLHRSGLTALSPH VTGLHRSIIANRSTYFLGLGSPSTVDSQSSLSVSHLACESLKRGESTIALAGV NINI VPESALGAAKFGGLSPDCRFTFDARANGYVRGEGGTIVLKLARAVADGDPV YCVRGALVNDGGDGLTVLQSGOEQVRLIAYRAGVNPADGVYELHGTGTRVGD PTEAAAGAVNGTGREPRLRVGSAKTVNHLEGAAGITGLLKAVLSLNRHLPASL NFGTNPATPLRLNLRVQTEHSEWAEDRPLLAGVSPFGMGNTNCHIVLASGAAVE AEGATAGGGAAPTWPISAKTASGLRAQAALLAATEAPGLALPVGWSLARTGTA FBRHVRVDRDDFLURALERSTGGIDAAITTRAGPRLGELALFDSGGSGRAGMR ELAESVPFASALDECAHLDFLLPRLPRFMFAPVGTGEAAELDQTLTYOTSPLAVE VALFRLESWDITPDVLMGHSIGELAAHVAGVLSADACALVAARGRLMPLPAGGA MIAVQAEDEPRMERTGERTDRVSI AALNGPDSVVSVDDELATEADAPAAALGRKTSR LRVSHAFHSPHEMPLAEFRNVAEGITFHA PIPVSNVTGLRSEPSGECTAAVY VRHREAVRPAQGVARLDQGVRTYILELPGGVGLASMARAGADSLSLFPALRTRPE AQALLAASLASHVGEVDPWDALFGGGRGARRRKVALPTVAFERQRWLDQTSVPAT TCWAPVAPQDAIAETAPVAGLRAGLAESEDEAVLDWVRAHTAAVILGHAEAR QVETWTFDLDGLDLSVSELNQLSEATGLRLPSGLLPDHPPTPAALARAQSLGIG GDAAAAATVADPHIEAI VAMSCRLPGCVSSPEDLWRLLASGGDGGISGFTDRGWD VETLVDPEPTGKSTRHGGFLHEAAEFDAQGFGLSPREAAIIPQORVLETAWEA FRAGIDPATLRSGRAGVIFGATHAQDYGRRLHEPADGVGGLTGTTSVAGSRVAYS FLEGPAVTVTDCSSSLVALHVAQSLRSGECTMALAGVTVMSTPGMFVBSFRQGR LADNGRCKAFAPAGDQAGVGGVLLVERLSDDRNRNGHVLA VVRGSAVNDQSGASG LTAPNGPSQORVVRQALAAAGLAPADVDAVEAHGTCTLLGLDPIEAELIAAYGQRDGG

CDS
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=1
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translation="MVSEELNATGNRDCSASADSADGIAVIGLCRLPGAADPAAPWR LLSREGADATDVPDWDGCAVADASAPDRGTGIRGGFLDSIGDFADGFFGLSPKEAA ANDPQORVLELAWALEDAHVRAATLSTRGTGVFGAMDDYATLHRSGLTALSPH VTGLHRSIIANRSTYFLGLGSPSTVDSQSSLSVSHLACESLKRGESTIALAGV NINI VPESALGAAKFGGLSPDCRFTFDARANGYVRGEGGTIVLKLARAVADGDPV YCVRGALVNDGGDGLTVLQSGOEQVRLIAYRAGVNPADGVYELHGTGTRVGD PTEAAAGAVNGTGREPRLRVGSAKTVNHLEGAAGITGLLKAVLSLNRHLPASL NFGTNPATPLRLNLRVQTEHSEWAEDRPLLAGVSPFGMGNTNCHIVLASGAAVE AEGATAGGGAAPTWPISAKTASGLRAQAALLAATEAPGLALPVGWSLARTGTA FBRHVRVDRDDFLURALERSTGGIDAAITTRAGPRLGELALFDSGGSGRAGMR ELAESVPFASALDECAHLDFLLPRLPRFMFAPVGTGEAAELDQTLTYOTSPLAVE VALFRLESWDITPDVLMGHSIGELAAHVAGVLSADACALVAARGRLMPLPAGGA MIAVQAEDEPRMERTGERTDRVSI AALNGPDSVVSVDDELATEADAPAAALGRKTSR LRVSHAFHSPHEMPLAEFRNVAEGITFHA PIPVSNVTGLRSEPSGECTAAVY VRHREAVRPAQGVARLDQGVRTYILELPGGVGLASMARAGADSLSLFPALRTRPE AQALLAASLASHVGEVDPWDALFGGGRGARRRKVALPTVAFERQRWLDQTSVPAT TCWAPVAPQDAIAETAPVAGLRAGLAESEDEAVLDWVRAHTAAVILGHAEAR QVETWTFDLDGLDLSVSELNQLSEATGLRLPSGLLPDHPPTPAALARAQSLGIG GDAAAAATVADPHIEAI VAMSCRLPGCVSSPEDLWRLLASGGDGGISGFTDRGWD VETLVDPEPTGKSTRHGGFLHEAAEFDAQGFGLSPREAAIIPQORVLETAWEA FRAGIDPATLRSGRAGVIFGATHAQDYGRRLHEPADGVGGLTGTTSVAGSRVAYS FLEGPAVTVTDCSSSLVALHVAQSLRSGECTMALAGVTVMSTPGMFVBSFRQGR LADNGRCKAFAPAGDQAGVGGVLLVERLSDDRNRNGHVLA VVRGSAVNDQSGASG LTAPNGPSQORVVRQALAAAGLAPADVDAVEAHGTCTLLGLDPIEAELIAAYGQRDGG

gene
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=1
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translation="MVSEELNATGNRDCSASADSADGIAVIGLCRLPGAADPAAPWR LLSREGADATDVPDWDGCAVADASAPDRGTGIRGGFLDSIGDFADGFFGLSPKEAA ANDPQORVLELAWALEDAHVRAATLSTRGTGVFGAMDDYATLHRSGLTALSPH VTGLHRSIIANRSTYFLGLGSPSTVDSQSSLSVSHLACESLKRGESTIALAGV NINI VPESALGAAKFGGLSPDCRFTFDARANGYVRGEGGTIVLKLARAVADGDPV YCVRGALVNDGGDGLTVLQSGOEQVRLIAYRAGVNPADGVYELHGTGTRVGD PTEAAAGAVNGTGREPRLRVGSAKTVNHLEGAAGITGLLKAVLSLNRHLPASL NFGTNPATPLRLNLRVQTEHSEWAEDRPLLAGVSPFGMGNTNCHIVLASGAAVE AEGATAGGGAAPTWPISAKTASGLRAQAALLAATEAPGLALPVGWSLARTGTA FBRHVRVDRDDFLURALERSTGGIDAAITTRAGPRLGELALFDSGGSGRAGMR ELAESVPFASALDECAHLDFLLPRLPRFMFAPVGTGEAAELDQTLTYOTSPLAVE VALFRLESWDITPDVLMGHSIGELAAHVAGVLSADACALVAARGRLMPLPAGGA MIAVQAEDEPRMERTGERTDRVSI AALNGPDSVVSVDDELATEADAPAAALGRKTSR LRVSHAFHSPHEMPLAEFRNVAEGITFHA PIPVSNVTGLRSEPSGECTAAVY VRHREAVRPAQGVARLDQGVRTYILELPGGVGLASMARAGADSLSLFPALRTRPE AQALLAASLASHVGEVDPWDALFGGGRGARRRKVALPTVAFERQRWLDQTSVPAT TCWAPVAPQDAIAETAPVAGLRAGLAESEDEAVLDWVRAHTAAVILGHAEAR QVETWTFDLDGLDLSVSELNQLSEATGLRLPSGLLPDHPPTPAALARAQSLGIG GDAAAAATVADPHIEAI VAMSCRLPGCVSSPEDLWRLLASGGDGGISGFTDRGWD VETLVDPEPTGKSTRHGGFLHEAAEFDAQGFGLSPREAAIIPQORVLETAWEA FRAGIDPATLRSGRAGVIFGATHAQDYGRRLHEPADGVGGLTGTTSVAGSRVAYS FLEGPAVTVTDCSSSLVALHVAQSLRSGECTMALAGVTVMSTPGMFVBSFRQGR LADNGRCKAFAPAGDQAGVGGVLLVERLSDDRNRNGHVLA VVRGSAVNDQSGASG LTAPNGPSQORVVRQALAAAGLAPADVDAVEAHGTCTLLGLDPIEAELIAAYGQRDGG

FLSSEIAEFQPNSTLPRPRAELDEPIALVGMACRPGVTSADDFWDLISSEQDAI
GGPTDRGMDLDTLDPDHPGTCYTRNGGLYDAGHFDADFPGISPREALAMDQQ
RLLETAWEIHEAGNPHHTLGTPTGVTGNGQDHAHIAIRQAPSEGFVLTGAAT
STASGRISVILGDEPAVLTACSSSVALHLCASLRSECTMALAGCATVMTPI
TTFEPARQGLADGRCKAFSAADGTGNGVGLLVERLSDARNRHRVLAUVTRI
ANVDGASNGLTAPNGPSQORVIRQALANADLTPADVAVEAHGTGTLTGDPIEAQAI
LYVQDGRFGLPLWGSVKSNGVHTQAAAGVAGVTKMVALRHRLTLPPLTHADEPSP
HWDMSAGVOLLITETVPFGEGRPRRVSFVSGTNAHVILEAPADDPVGGPPA
DEADGSEBAALAGSDPQWPLWSAKSPALRAQAQALHAHLTDHPGLDIADLYTLAH
ARAVFHRATLAADRDFTLOALQAAGEPHAPVTHSSAPGCTGGEAGAKTATCS
GQOTRPGMAHGLYHVPFAALANDICTHLDPLHDPILLPLTQNDNDNEDAAAL
LAQTPAQPALFAQVALHRLTLDGHIHTPHYAGHSIGEITAAHLAHLITDATTTL
ITQATLMDMPGTTLHTTHHLLTAHENDLAAINTPISLVISPTHTVQ
HITLTCQOQGIKTPLTNAHSPHTNPLNQLHQHTQTLTTPHPTPLTANTPPD
QLLTHYMTQOARNTVDYATTTQTLHQGVTVYIELGPONTLLTTHNLNPTPTT
TUTHPHHPQHLNLAKTTTTHPHVHTHNOPTHHTLDDLTPTFQHHYHLEL
PSAOTSGORRSPRSDPTAESFEDVANEEDLSLAETLDDASALDVPVLSNHW
RHQDOARINTWYQETWKLPLTPHQPQHLWLAIPETQTHHPTNLTNLDNHHW
ITPIPLVTNHTNPOHLHTLHTRQOQNHHTGPIITGLLSLLALDETPHHPHTP
TGTLLNLTPQHTPTPLWYATNATTHPNPDPLTHTPTQAOTWGLARTLLLEHP
THTAGIIDLPTPTHTLHTLTHLTQHTPHQTLALRTTGTHTRLTPTLTPTOPP
TTPHTGTLITGTAALHTLHTLTHLTQHTPHQTLALRTTGTHTRLTPTLTPTOPP
IHLTITCDSNPDLQOLNLTIPPOHPLTTHVHTAGVNLFPVSETDAESVTA
KATGAAILHELLDHTLEHPIFLSSGAGWGNOCAYSAANAYLDALATHROTHGL
PGASAWGAGKSGADAHGYLEKRGILPMEPRMALAAPHRAQPNLSNLIAD
IDWERFPAFTARRHSPLIEDIEVRQAQEELEAASTAKTTTAQPIATLSRLERLAR
TSSKQNVLLGIIRITGICVLGRPEGEIDORAFDLGDSLTQAQFSKELAKETGL
PLPPLSVFPYPTPOEAAHLTQVLDLDEDAALSNALPOVAHRTVEDEPIAIIIGM
ACRFPGVPSADLLMELASGKDAIGVPTDRGMDLTDLPDPHPGTCYTRNGGL
YAGHFDABFGISPREALAMPQRLLETAWETIEHAGINPHLHTPTGTVFAGIN
AODHAHRIHQSHDVTIEGYALTGSSGVSASGRVAYTLGLEPAVSVDPACSSSILVAL

Query Match 72.6%; Score 944.8; DB 1; Length 64957;
Best Local Similarity 83.6%; Pred. No. 1.9e-110;
Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;

QY 1 GACCCGATCCCATCATCGGCATGCGATCCGTTTCCCGGCGGAGTCGCGTCCCGAAG 60
DB 49923 GAGCCGATTGCGATTGCGGGATGCGTGTCTGTTTCCGGGTGGGTGTGTTCCGCGGAG 49864

QY 61 GACCTGTGGAGCTGCGCCCTCGGCGGAGAGCGCATCGGCGCGTTCCTCCACGACCGC 120
DB 49863 GAGTTGTGGAGCTGTTGCTGCGGTGGGATGCGATTGGTGAATTTCCCGCCCGGTGCG 49804

QY 121 GGATGCCCCACGAAACAGCGTCACGCCACGACCCACGACGCCCGGACGTTCTATCCG 180
DB 49803 GGTGGGATCTGAGGGGTGTTGATTCCGACCTGACCGGTGCGGAGCGTCTGACGC 49744

QY 181 CAGGAGGCGGTTCCTTCCAGACGCGCGCATCTTCAGCCCGCGTCTTCGGAATCACT 240
DB 49743 CGGTATGCGCGGTTTTTGTATGAGCGGGGAGTTCGATGCGGACTTCTTCGGGATCACT 49684

QY 241 CCACGTGAGGCACTGGCGATGATCCGAGACGCGCTCTCTGAGACGTCCTGGGAG 300
DB 49683 CCGCGTGAAGCGGTGGCGATGATCCGACGACGCGTGTGCTGAGACGTCGTGGGAG 49624

QY 301 GCGTTGAGCGCGGGAATCGATCCGCTGTGCTGACGCGGTCCGCTACGCGCGTCTTC 360
DB 49623 GCGTTGAGCGCGGCGGTATCGATCCGCTGTGATGCGGTGGCTCCGCTACGCGGTCTTC 49564

QY 361 GCGGCGCGCTCTCTCTTCGATACGCGCCCGCTATGGACACCGCGTCTGCGAGGCGCGC 420
DB 49563 GCGGCGGTGATGTACACGACTACGATCCCGCTGGGTACCAT-----CCCC 49516

QY 421 GCGGAGCTGGAGGCGCATCTTACCGGTACACGCGGAGCGTCTCTGTCGGGCGGTATC 480
DB 49515 GAGGATTCGAGGCGGTATCATCGGCAACGATAGCGCGCGCGCTCGCGTGGCGCGCGTC 49456

QY 481 GCCTACAGCTTCGGGCTGGAAGCGCGGCGATCACCGTGGACACGCGGTCTCGCATCG 540
DB 49455 GCCTACAGCTTCGGTCTGAGGGCGCTGCGCTCTCGTGGACACGCGCATGTTCTGCTG 49396

QY 541 CTCGTGACGCTGCATCTGCGCTGCCAGTCTGCTGCGGTGCGGTGAGTGACGCTGCGGCTG 600

DB 49395 TTGTTGGCGCTGCATCTGCGCTGCCAGTCCCTGCGGTGCGGTGAGTGACGCTCGCGCTG 49336
QY 601 GCGCGCGCGCTCTCGGTCTCATGCTCCACCTCGGCATGTTTCATCGAGTCTTCTCCCGCGACGCG 660
DB 49335 GCGCGCGGTGTGACGGTGTGATGTCGACCCCGCACTCTTCTGTCGAGTCTTCACGCCACGCG 49276
QY 661 GGGCTGTTCGGTGGACGCGAGGTGCAAGGCGTACTTCGGCTGCGACCGACCGCACCGCGTGG 720
DB 49275 GGACTGTTCGGTGGACGCGCGCTGCAAGTCTTTCGCGGGTGGAGCCGACCGCACCGCATG 49216
QY 721 GCGAGGCGCTCGGATGCTGTTGTTGTTGAGCGGTGTTGCGATGCGGTGCGGTGCGGTCAT 780
DB 49215 GCGAGGCGCTCGGATGCTGTTGTTGTTGAGCGGTGTTGCGATGCGGTGCGGTGCGGTCAT 49156
QY 781 CGGCTGTTCGGCGGTGTGTCGCGCGAGTGCAGCGCGGTCAACACGACGCTGCGTTCGATGGGCTG 840
DB 49155 CGGCTGTTCGGCGGTGTGTCGCGCGAGTGCAGCGCGGTCAATCAGACGCTGCGTTCGATGGGTTG 49096
QY 841 ACGGCGCCGAAACGCTCCGCTCAGGAGCGCGGTGATCCGCGACGCGCTTGGCGAACGCGGGG 900
DB 49095 ACGGCGCCGAAATGGTTCGCTCAGGAGCGGTGATCCGCGACGCGGTGTTGGCGAACGCGGGG 49036
QY 901 TTGTCGTGCGCGATGTCGATGTTGTTGAGGCGCGACGCGACGCGCTGCGTTCGAT 960
DB 49035 TTGTCGTGCGCGATGTCGATGTTGTTGAGGCGCGATGCGCGACGCGCTGCGTTCGAT 48976
QY 961 CGGATCGAGGACGCGGTGTCCTCCACGTCAGGCGCGCGCGCGGTGTCACAGCGCGCTG 1020
DB 48975 CGGATCGAGGCGCGCGGTGTCCTCCACGTCAGGCGCGCGCGCGGTGTCACAGCGCGCTG 48916
QY 1021 TGGCTGGGCTCTCTGAAGTCCAAACATCGGCGCACACATCGCTGCGCGCGGTGTTGGGTGG 1080
DB 48915 TGGCTGGGATCGTGTGATCGAACATCGGCGCATCGCGAGCTGTCGCGCGGTGTTGGGTGG 48856
QY 1081 GTTCATCAAGATGGTGTGATGCTGTTGCGGAGGCGGTGTTGCCCGGACGTTTGCATGTTGAT 1140
DB 48855 GTTCATCAAGATGGTGTGATGCTGTTGCGGAGGCGGTGTTGCCCGGACGTTTGCATGTTGAT 48796
QY 1141 AAGCGCTGCGCGCGAGTGGATGCTGTCGCGCGCGCGGTGCGGCTGCTGACGAGGCGGTG 1200
DB 48795 GAGCGCTGCGCGCGAGTGGATGCTGTCGCGCGCGCGGTGCGGCTGCTGACGAGGCGGTG 48736
QY 1201 CCGTGGCGCGGCGACGCGCGCGCGGTGTCGCGCGCGCGGAGTGTCTGCTTTCGCGATC 1260
DB 48735 CCGTGGCGCGGCGACGCGCGCGCGGTGTCGCGCGCGGAGTGTCTGCTTTCGCGATC 48676
QY 1261 GCGCGCACGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 1302
DB 48675 AGTGCGCGAATGCGCATGTGATTTTGGAGGAGGCGCGCGCG 48634

RESULT 7
AB070949
LOCUS Streptomyces avermitilis 78210 bp DNA linear BCT 11-OCT-2001
DEFINITION Streptomyces avermitilis polyene macrolide biosynthetic gene cluster.
ACCESSION AB070949
VERSION AB070949.1 GI:15824136
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis DNA.
Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403

Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikedam@pharm.kitasato-u.ac.jp. Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)
Location/Qualifiers

1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
101. .12019
/gene="aveA1"
101. .12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDCGEPRPAAEVLGVADBADGGVVFVFGQGPQWPMGR
ELDASDFVESVRAECAFAPFYVDNSVEQLRSDPAGLDRVVDVQPTLFLAVMSTL
AAWRGSGVPCFALHSGEIAAAHVSGLSLADAAAVTLMSQAVTLFAGTICALVSV
VAATPDELLPRIAPWEDNPARLAAVANGPRSTVSVSGAEAAVLADLTAAQVYTR
MIPVDPAHSPLYAIEERVVSGLLPTTPRSPRIPIHSSSVTGGRLDRELDAAVYTR
MSSTVPEAPARLLDQGPFTVEMSPHPTVMGLQELADLGGDTGTADTVINGTLR
DGNRTDHFSTLSLAQRHGETSATVLSARLTAQSTQOSSLDDIRHARTAAVNDLR
KNGERTASDAGPASFALGFSVMGVELRNLSKATGLRPTVLI PDHPTTAAVARD
LRTAALGHDDTPVPSGHHGTAAADPTAIGWACRFPGVSRPKDLMELAAS
GGDAI GPHEDDTPVPSGHHGTAAADPTAIGWACRFPGVSRPKDLMELAAS
MDPQRLLETSWEAFRAGIDPLSVRSGRTGVPAGALSFDYGPMDTASSEGNAUVE
GHILTGTGVSIGRISVYSGLEBPAITVDTGCSASLTLLHACQSLRSECTLALAG
GVSMTSLGFI EFISROGLSVDCRKAYSAAADGTGMEGVMLVERLSDAVRLGG
RVLA VGRGSVANDGASGLTAPNGPAQVRIQALANAGLSVADVDVVEGHTGTTL
GDPTAEQAALLTGORAGDPLVGLSKNIGHTMAAGVGVTKMVALREGVLPRT
LHVDKSPQVDWSAGVRLTEAVPWGDDAGLRAGVASFGIGGTNAHVILBEAPAA
AGCGVAGGVELEGAPLIVSAESVAAPVASVAVSVVPPVPPVPPVPSVSRSEAG
LRAQEAELRVAVPVDSIADVAGLACGRAVLEHRAVLAAADRELVOGLGALAAG
EPDRVTTHGAPGDGGVVFVFGQGPQWGMGRVLLASVPFARMQACEALAPSW
VDSNVDDI LRRDAGDAWEADVVQVPLFVMSLAALWRSYGI EPDAVLHSGOETA
AAHVCGALS LKDAKTAVALSRALAAVGRGGMASVLPAAQEVQLIGERWAGRLWA
ANVGPRSTASGDAAVEVLVACAGTGVARRIPVDYASHCHVQVLEELLEGD
ISQPSGTVFVEGTVLDTTLDAAAYWRNLHQPVRFSDAQLAADGHRVFEVSEV
THHTHPVAIEDTDEADTVHIGLSRRGNDTRFLTA LAHTHTTGITGTPTWHHHY
TVHHTHPVAIEDTDEADTVHIGLSRRGNDTRFLTA LAHTHTTGITGTPTWHHHY
EVESRFDAVAROQDETVATLAVPPSAGLTVPVPAALSAWRHQQHQAQNTWYQST
WKPLTLPTTHQPHOETVLAIPETOTHTHPIITLNLNHHGIIPIPLTLANHTNPOH
LHHTLHTRQAQNHGTGATGLLSLALDETPHPHPTPTGTLNLTLTQTHTQTH
PPTPLWATTNATHTNDPLHTPTQATGLARTLLEHPTHTGATDILPTLTGTTGAL
LQHLTOTLPHQTOQLARTGTHTRRLTPTLTPHTPTPTGHTLTLTGTTGAL
ALNLTHTLTHQPTQHLLTSTRGTHPTPHQTLTQQQGIHLITCTDTSNPDLQ
QNLNTI PPHQLTIVITAGLIDBATLNTLTQNLNLRKASHLHLOHTHTLPT
TAFVLVSSAATFCAGQANVAAAVLDALAHRTTHLIPATSIANGTQWGNGLADS
DAKRAYLDGRFRFSPMLATAVQAIADTERFVVIADIDNSKI EHTSQSTDLVSA
AREREPADRPPEALHQLAHTQADQRAALLERHVDHVAALRHADPKAIAPDQDS
FRALGDSVTEFERNLTKATGLRPLSVLVEPTPAKLAVHLQNGRLGTGAASAPFS
AAAVTASVTEPTAIVGMACRFPGVTSADDFWDLISSEODATLQPTDRGMDLTPL
YDPPDHGTCTGTNGGLVDAGHDEAFEGGISPREALANDPQORLLELTAWETIEHA
GINPHTLHGTPGTGTNGQDYALRVHNAQSGDGFALGTGASVTSGRISYTFGFE
GPVSDTVLACSSSLVALHCAQALRAGECSMALAGGVTVMSSPGAFVFSQRLGAAD
GHCKFAAAGADGTGMEGVMLVERLSDAHNRGHVLA VVGRSADVNDGASGLTAP
NGPSQQRVIRQALANAGLSAGDVAVEAHGTGTTLGDPIEAQALLATVQDRAEGEPL
WLGSGNVGHTQAAAGVGIKVMALRHGLLPRTLHVDSPSPHVDKASACAVOLLTE
TPWPGGEGRLRRAVSGVSGVSNHVILBEAPADVDVPGGPAGGAGSDGADRAAAG
SPGVWPLVLSAKSPALRAQAQALHAHLTHPHGLDLADVGYTLAHARAVPHRATLNA
ADRTFLQALAAAGDEPHPAVHSSAPGGTGEAGKTAFCSSQGTQRPQWMAHGL
YHTHPVFAALANDICTHLDPHLHPLPLTLDNDNDNDAALQOQRYAQAPALFQ
VALHRLTDGYHTIPTHYAHENDLSGEITAAHLAGLITLTDATTLTQATRLMTPPQPT
MTTLHTTPHITTHLTAYENDLSGEITAAHLAGLITLTDATTLTQATRLMTPPQPT
LPTNHA FHSPTNLNQLHOHTOTLVHPHPTPLIANTPPDOLLPHYWTQOARNTL
VDYATTTQHQGVTVIELGPNOTLTTLTHNLNPNPTTLTLTHHPHQTHLLT
NLAKYTTTWHPHHTHNDQPHHTHLLDPTYPFHQHHYWLSETPQAGNVSAAGDL
TEHPFLGALATDGGALLAGRLSRHSPWLADHAGVGTLLSGATLELALHAGTV
VGCPRVDELTLHAPVLDVGVSGVGVAAADGEGRLSVVYARGSGACSGGASGSG
VGCPRVDELTLHAPVLDVGVSGVGVAAADGEGRLSVVYARGSGACSGGASGSG

SOURCE Streptomyces avermitilis.
 ORGANISM Streptomyces avermitilis
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 1 to 31422)
 AUTHORS Omura, S. and Ikeda, H.
 TITLE Avermectin aglycon synthase gene
 JOURNAL Patent: JP 2000245457-A 2 12-SEP-2000;
 THE KITASATO INSTITUTE
 COMMENT OS Streptomyces avermitilis
 PN JP 2000245457-A/2
 PD 12-SEP-2000
 PF 24-FEB-1999 JP 1999046961
 PR SATOSHI OMURA, HARUO IKEDA
 PI C12N15/00 A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
 C12Q1/68//C07D493/22, (C12N9/88, C12R1:465), C12N15/00 CC
 PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
 FH Key Location/Qualifiers
 FT CDS (1)..(14643)
 FT CDS (14824)..(31419).
 FEATURES
 source
 1..31422
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 BASE COUNT 4136 a 10237 c 11677 g 5372 t
 ORIGIN
 Query Match 72.6%; Score 944.8; DB 6; Length 31422;
 Best Local Similarity 83.8%; Pred. No. 2.3e-110;
 Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;
 1 GACCCGATCCCATCATCGCATGCGATGCGTTCCTCCGGCGGAGTCGGTCCCGAAG 60
 14935 GAGCCGATTGCGATTGTGGGATGGCGTGTCTTCCGGTGGGGTGTTCGCGGAG 14994
 61 GACCTGTGGAGAGTGCCGCCCTCGGGCGGAGACGCATCGGGCCGTTCCCGACCGACGC 120
 14995 GAGTTGTGGAGCTGGTTCGCTCGGTGGGATGCGATTGGTGAATTTCCGGCCGGTCCG 15054
 121 GATGGCCACGACAGCGTCACGCCAGGACCCACGACCGACCGCGGACGTTCTATCCG 180
 15055 GGGTGGGATCTGGAGGGGTGTTGATTCCGACCCCTGACCGGTGGGGACGTCGTACGCG 15114
 181 CAGGAGAGCGGGTTCCTTCACGACCGGGCGCTTCGACCGCGGCTTCTCGGAATCAGT 240
 15115 CGGTATGGCGGTTTGTATGAGCGGGGGATTCGATCGCGATCTCTTCGGATCAGT 15174
 241 CCACGTGAGGACTGGCGATGATCCGACGAGCGGCTGCTCTGGAGACGTCCTGGGAG 300
 15175 CCGCGTGAAGCGTTGGCGATGATCCGACGAGCGGTTGTTCTGGAGAGCTCGTGGAG 15234
 301 GCGTTGAGCGGGGGGATCATCGCTGTCGATACGGGTCGCGGTCCGTCAGGGCGTCTTC 360
 15235 GCGTTGAGCGGGGGGATCATCGCTGTCGATACGGTGGTGGTCCGTCAGGGGTCTCTTC 15294
 361 GCGGCGCCCTCTCTTCGACTACGCGCGCGGTATGGACACCGCTCGTCGAGGGCGCC 420
 15295 GCGGGGTGATGATACACGACTACGATCCCGCTCCCGCTGGGTACCAT-----CCCC 15342
 421 CGGACGTTGAGGGGACATCTCACCAGTACACGCGGAGCGTCTCTCGGGCGCGTATC 480
 15343 GAGGATTCGAGGGCTACATCGGCAACGATAGCGCGGCGGCTCGCTCGGGCGCGCTC 15402
 481 GCCTACAGTTCCGGCTGGAAGGGCGGGGATCACCGTGGACACGGGTCGTCGCGATCG 540
 15403 GCCTACAGCTCGGTCTCGAGGGCCCTGCGCTCTCGGTGGACACGCGATGTTCTGTCGTCG 15462
 541 CTCGTGACGTCATCTGCGGTGCGAGTCGCTCGGTGCGGTGGGTGAGTCAAGCTCGCGCTG 600
 15463 TTGTGGCGGCTGCATCTGGCGTGCAGTCGCTGCCGTGGGTGAGTCAAGCTCGCGCTG 15522
 601 GCCGCGGCGTCTCGGTGATGTCACCCCTCGGCATGTTTCATCGAGTTCTCCCGGACGCGC 660

Db 15523 GCCGCGGTGTGACGGTGTGCGACCCCGACCTCTTCGTGAGTTCTTACGCCAGCGC 15582
 Qy 661 GGGCTGTGGTGGACGGCAGGTGCAAGGGTACTCGGTGCGAGCGACCGACCGCGCTCG 720
 Db 15583 GGACTGTGGTGGACGGCGCGTGCAGGTCCTTCGCGGGGTGGAGCGACCGGCACCGCATG 15642
 Qy 721 GCGGAGGGCGTCCGGATGCTGTGTGAGCGGTTGTGCGATGCGGTGCGGTGCGGTGGGCAT 780
 Db 15643 GCGGAGGGCGTCCGGATGCTGTGTGAGCGGTTGTGCGATGCGGTGCGGTGCGGTGGGCAT 15702
 Qy 781 CCGGTGCTGGCGGTGCTACCGCGCATGCGGTCAACCGAGACGCTGCGTCAATGGGCTG 840
 Db 15703 CCGGTGCTGGCGGTGCTGCGCGCATGCGGTCAATCAGGACGCTGCGTCAATGGGCTG 15762
 Qy 841 AGCGCGCGCAACGGTCCGGCTCAGGAGCGGGTGTATCCGCGAGCGGCTGTGGCGAACCGGG 900
 Db 15763 AGCGCGCGCAATGGTCCGGCTCAGGAGCGGGTGTATCCGCGAGCGGCTGTGGCGAACCGGG 15822
 Qy 901 TTGTCCGTGGCGGATGTGATGTGTGAGGGGACGGGACGGGACGACGACGCTGGGTGAT 960
 Db 15823 TTGTCCGTGGCGGATGTGATGTGTGAGGGGACGAGGGGACGACGACGCTGGGTGAT 15882
 Qy 961 CGATCGAGGACAGCGGCTTGTCTCGCCACGTAACCGGACGCGGCGCGGTCAACGCGCTG 1020
 Db 15883 CGATCGAGGACAGCGGCTTGTCTCGCCACGTAACCGGACGCGGCGCGGTAAACGCGCTG 15942
 Qy 1021 TGGTGGGGTCTCTGAAGTCCAAATCCGCGACACCATGGTTCGCGGGGTGTGGGTGG 1080
 Db 15943 TGGTGGGGTCTCTGAAGTCCAAATCCGCGACACCATGGTTCGCGGGGTGTGGGTGG 16002
 Qy 1081 GTCAATCAAGATGTGATGCGGTTCGCGAGGGGGTGTTCGCGGACGCTTCGATGTGGAT 1140
 Db 16003 GTCAATCAAGATGTGATGCGGTTCGCGAGGGGGTGTTCGCGGACGCTTCGATGTGGAT 16062
 Qy 1141 AAGCGTCCGCGAGGTGGACTGCTCCGCGGGGGCGGTTCGCGGCTCTCACGAGGCGGTG 1200
 Db 16063 GAGCGTCCGCGAGGTGGACTGCTCCGCGGGGGCGGTTCGCGGCTCTCACGAGGCGGTG 16122
 Qy 1201 CCGTGGCGGGGACGCGGACGGGGTTCGCGGGGGGGAGTGTCTGTTTGGGATC 1260
 Db 16123 CCGTGGCGGGGACGCGGACGGGGTTCGCGGGGGGGAGTGTCTGTTTGGGATC 16182
 Qy 1261 GCGCGCAGTAATGCGCATGTGATTTTGAGGAGGCGCGCGC 1302
 Db 16183 AGTGGCAGTAATGCGCATGTGATTTTGAGGAGGCGCGCGC 16224

RESULT 6

AB032367/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

AB032367 64957 bp DNA linear BCT 18-SEP-1999
 Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
 aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.

AB032367.1 GI:5902890
 AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
 4; type I polyketide synthase AVES 3; type I polyketide synthase

AVES 2; type I polyketide synthase AVES 1.
 Streptomyces avermitilis DNA.
 Streptomyces avermitilis

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 64957)

Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S.

Organization of the biosynthetic gene cluster for the polyketide

anthelmintic macroide avermectin in Streptomyces avermitilis

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)

99380548

2 (bases 1 to 64957)

Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S.

Direct Submission

Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical

Sciences, Kitasato University, Microbial chemistry; 5-9-1

AKTVALPQVEBQLIGERGLWVAANGPRSTAVSGDAEAVVEVLAYCAGTGVPRPI
PVDYASHQVQALBELLELLDSDISPOQYVFPFSTVEGTWLDITDLDAAVYNLH
QVRSFSDHVALADQGHVLEVSHPHTLPVPAIEDTREDTARRHCDQPPRRRHP
LLPQLRLDPYRHRQTHVAPLDDPPHLPPLDAPPAHLPFOHQHYWLESSQGA
GGSGAGSGAGSAGTAGTAEVSRFDVARQDLETAIVATLPPSAGLDTV
PALSARHQRHQDARINTWYQETWKLPTLTPHQPHQTLIAIPEYTHPHHTNII
NLHHHTITPIPLPTHTHTNPHQHHTLHTHQQAQNHHTGATIGLLSLALDTPH
PHHPHTPTGTLNLNLTOHTHTPTPLWYATNTATTHPNDPLTHPTQAQWGLAR
TLLBHPHTTAGIIDLPTPTPHLHLTOTLTPHQHOTOLAIRTTGTHRELTPTL
TTPHOPPTPHGTTLTGCGTALATHLTHLTTHQPHQLLLTSRTGPHTPHAQHLT
TOLQNGHLITITCOTSTPRTHNSLNTIPPHPTTVIHTGLDOLATLNTIPT
QNNVLRKAKHSAHLLQHTPLTAFVLYSSAAATGAPQOANYAANAYLDALH
KHHTHLPTSATWGTNOGLADSKARAYLDRRGPMPSPELATAVQAIADTER
PVYADIDWSKIEHTSOTDLVSAAAREPAPQRTPEALHKTLAHOTSQDAAL
LELVADHVAALRHADPKAIPDOQFRALGFOSLTAVBFRNLLIKATGLRLPVSLVFD
HPTAKLAHLQNLQRTGAESAAPSAAVTAESVTEPIAVGMACRPPGVTSADDF
WDLISSEQDAIGGFTDRGNDLDTLYDDPPDHPTCYTRNGGFLDAGHDFABPFGIS
PREALMDDQALLLETAWETIEHAGINPHHTLHPTGTVFTNGODYALRVHNAQS
TDGFALTAGSVISGRISYTFEGEPVAVSDTACSSSLVALHLACQALRAGECSMAL
AGCVTVMSPGAFVFSRQGLAHDCHKAFSAADGTGMEGVGMGLVERLSDAHN
GHRVLAVVSGAVNODGASNGLTPNGPSQORVIQOALANAGLSAGDVDAEHAQGT
TIGDPTLEAQLATYGOPRAGGELWLSGVKSNGVHTQAAAGVAVIKWVALRGLL
PRTLVHRSRPHVDSAGAVQLTETVPHPGEGELRRAGVSSFGVSGTNAHVILEEA
PADDPVPGPAGGEGAGSDDEAAGSPGVPWMLVSQOPALRAQAPAVIHSSAPGCTG
LDAVDVGYTLAHARAVDFHRAATLAAEDRTFLQALAAAGPAPAVIHSSAPGCTG
GEAAKTAFCISGGQTORPMAHGLYHTHPFAAALNDICTHLDPHLDHPLPLLTQN
DNDDEAALLQOTRYAQPALFAVALHRLITDGVHITPHYVAGHSLEITGAHLAG
ILTDATLTLTQALACTMPTGTTLHTPHHTLHTAHENDLALAIANTPTSL
VLSGPHTVQHTITCQQQIKITLPTNHAFSPHTNFIQLNQHQAOTLTYHPHPT
PLITANTPQDLTTHYTOQARTVDYATTTOTLHOHQVITYIELGPDNTLTTLTHH
NLPTPTTLTLTHPHHPHTLNLAKTTLTTHPHYTHHAQPTPHPLDLPSTYP
FOHHYVLESTOPGAGNVSAGLDPTEPLLGATLELALEADTGALLAGLSLASHPLA
DHAGVTLLSGATFLELALHAGTVGCDRVDLTLHAPLVVDPGVSVQVYVAAAD
GHRBLSVYVAGSGSAGGSGVGTCHASGLVBAAGSVVDVGLAGVWPPRGAV
AVDVGDRRLAGACVLPVFSGLAVVRDGDLLAEVCLPEBAWGDAGLPLPAL
LNGVQPLSVLLPGTGFKGAGFGVGRVPAVWGVSLHRAVGTGVVRVSAVGRG
GREAVSVVSDAGVPAVSVDRELEPVDMLQRLAVSVSAGRRLSLYAVQWAEVGPV
VCCQAWMHEDVGGSGGVPVGVVLRCPDAGAGGGGGGGGGVGVGVVGVVQ
WGLGRFASRLVTVTRGAVAPEDGPVDVGVASVWGLVRSQAQHPDRFVLLDLD
DTGTDLTGAGAGWGDGGRVAVVACGEPOLAVRGERLLAARLTFLRESSGDVPAQRS
GUTRARSDVPAQRSQGVPAKRSVDVSGREVLPWLSGGSVLTGTGTGVIGAAVAKHLA
GVGVDRLLLVSRRGDAPAGLRSRWPRAEVRIVACDVGERRVERVRLAGGCSRV
SVFVDFSAVAVTVPLPVASDVKKAAMAYATVEPTDYLDPPD"

BASE COUNT 1756 a 4059 c 3528 g 1753 t

Query Match 86.7%; Score 1128.8; DB 1; Length 11096;

Best Local Similarity 94.2%; Pred. No. 2.2e-133;

Matches 1226; Conservative 0; Mismatches 22; Indels 54; Gaps 3;

QY 1 GACCCGATCGCCATATCGGCATGCGTTCCTCCGGGGGAGTCCGGTCCCGGAAG 60
DB 1603 GACCCGATCGCCATATCGGCATGCGTTCCTCCGGGGGAGTCCGGTCCCGGAAG 1662
QY 61 GACCTGTGGAGCTGCGCTCGGCGGAGACGCCATCGGCGGTTCCTCCACCGACCGC 120
DB 1663 GACCTGTGGAGCTGCGCGCTCGGCGGAGACGCCATCGGCGGTTCCTCCACCGACCGC 1722
QY 121 GGATGCCCCACGACAGCTACGCCCGAGGACCCACCGAGCCCGGACAGTTCATCCG 180
DB 1723 GGATGCCCCACGACAGCTACGCCCGAGGACCCACCGAGCCCGGACAGTTCATCCG 1782
QY 181 CAGGAGGGGGTTCCTTACAGACGGCGGCACTTCGACGCGGCTTCCTCGGAATCAGT 240
DB 1783 CAGGAGGGGGTTCCTTACAGACGGCGGCACTTCGACGCGGCTTCCTCGGAATCAGT 1842
QY 241 CCACGTGAGGCACTGCGCATGGATCCGACAGCGGCTGCTGCTGGAGACGCTCTGGGAG 300
DB 1843 CCACGTGAGGCACTGCGCATGGATCCGACAGCGGCTGCTGCTGGAGACGCTCTGGGAG 1902
QY 301 GCGTTGAGGGGGGGAATCGATCCGCTGTGCTGACGGGGTCCCGTACGGGCGTCTTC 360
DB 1903 GCGTTGAGGGGGGGAATCGATCCGCTGTGCTGACGGGGTCCCGTACGGGCGTCTTC 1962

RESULT 5

E38021

LOCUS

DEFINITION Avermectin aglycon synthase gene.

ACCESSION E38021

VERSION E38021.1

KEYWORDS JP 2000245457-A/2.

E38021

31422 bp

DNA

linear

PAT 31-JAN-2002

```
QY 1 GACCCGATCGCCATCATCGGCATGCGATCCGCTTCCCGGGCGGAGTCCCGTCCCGGAAG 60
Db 2258 GACCCGATCGCCATCATCGGCATGCGATCCGCTTCCCGGGCGGAGTCCCGTCCCGGAAG 2317
QY 61 GACCTGTGGAGTGGCGCTCGGCGGAGAGCGCCATCGGGCGGTTCGCCACCGACCGC 120
Db 2318 GACCTGTGGAGTGGCGCTCGGCGGAGAGCGCCATCGGGCGGTTCGCCACCGACCGC 2377
QY 121 GGATGCCACCGAAGACAGCGTCAACGCCAGGACCCACAGCAGCGCGGACGTTCTATCCG 180
Db 2378 GGATGCCACCGAAGACAGCGTCAACGCCAGGACCCACAGCAGCGCGGACGTTCTATCCG 2437
QY 181 CAGGGAGGCGGTTCTTCAACGACGCGGCGCATCTTCGACGCCGCGTCTTTCGGAATCAGT 240
Db 2438 CAGGGAGGCGGTTCTTCAACGACGCGGCGCATCTTCGACGCCGCGTCTTTCGGAATCAGT 2497
QY 241 CCACGTGAGGCACTGCGCATGATCCGACGAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
Db 2498 CCACGTGAGGCACTGCGCATGATCCGACGAGCGGCTGCTGCTGGAGACGTCCTGGGAG 2557
QY 301 GCGTTCGACGCGCGGGAATCGATCCGCTGTCGCTACGCGGTCGCGTACGCGGCGTCTTC 360
Db 2558 GCGTTCGACGCGCGGGAATCGATCCGCTGTCGCTACGCGGTCGCGTACGCGGCGTCTTC 2617
QY 361 GCGGGCGCCCTCTCTTCCGACTACGCGCCGCGTATGGACACCGCGCTGCTCGAGGCGCGC 420
Db 2618 GCGGGCGCCCTCTCTTCCGACTACGCGCCGCGTATGGACACCGCGCTGCTCGAGGCGCGC 2677
QY 421 GCGGAGTGGAGGCGCACATCTCTACCGGTACACGCGGAGCGTCTGTCGGGCGGTATC 480
Db 2678 GCGGAGTGGAGGCGCACATCTCTACCGGTACACGCGGAGCGTCTGTCGGGCGGTATC 2737
QY 481 GCCTACAGCTTCGGGCTGAAGCGCGGATCACCGTGGACACCGGGTCTCGGCATCG 540
Db 2738 GCCTACAGCTTCGGGCTGAAGCGCGGATCACCGTGGACACCGGGTCTCGGCATCG 2797
QY 541 CTCGTGACGCTGATCTGCGGTGCCAGTCTGCGGTGCGGTGAGTCAGTCACGCTCGCGGTG 600
Db 2798 CTCGTGACGCTGATCTGCGGTGCCAGTCTGCGGTGCGGTGAGTCAGTCACGCTCGCGGTG 2857
QY 601 GCGGCGGCGTCTCGGTATGCTCCACCTCGGATGTTTCATCGAGTCTTCCCGGAGCGC 660
Db 2858 GCGGCGGCGTCTCGGTATGCTCCACCTCGGATGTTTCATCGAGTCTTCCCGGAGCGC 2917
QY 661 GCGCTGTGCGTGCAGCGGAGTGCAGGCGTACTCGGCTGCAGCGGACCGGACCGGCTGG 720
Db 2918 GCGCTGTGCGTGCAGCGGAGTGCAGGCGTACTCGGCTGCAGCGGACCGGACCGGCTGG 2977
QY 721 GCGGAGGCGTCTCGGATGCTGTTGTTGGAGCGGTTGTCGGATGCGGTCGGTGGGGCAT 780
Db 2978 GCGGAGGCGTCTCGGATGCTGTTGTTGGAGCGGTTGTCGGATGCGGTCGGTGGGGCAT 3037
QY 781 CCGGCTGTCGGCGGTGATCGCGGATGCGGTCAACCGAGGACGGTTCGATCGAATCGGCTG 840
Db 3038 CCGGCTGTCGGCGGTGATCGCGGATGCGGTCAACCGAGGACGGTTCGATCGAATCGGCTG 3097
QY 841 ACGGCGCGAAGCGTCCGCTCAGGACGGGTGATCCGCGAGCGGTTGCGGACCGCGGG 900
Db 3098 ACGGCGCGAAGCGTCCGCTCAGGACGGGTGATCCGCGAGCGGTTGCGGACCGCGGG 3157
QY 901 TTGTCGTCGGCGGATGTCGATGTCGTTGGAGGGGACCGGACCGGACCGTCGTCGAT 960
Db 3158 TTGTCGTCGGCGGATGTCGATGTCGTTGGAGGGGACCGGACCGGACCGTCGTCGAT 3217
QY 961 CCGATCGAGGACACGCGTTCCTCGCACGTCGAGGACCGGCGCGGTGACAGCGCGCTG 1020
Db 3218 CCGATCGAGGACACGCGTTCCTCGCACGTCGAGGACCGGCGCGGTGACAGCGCGCTG 3277
QY 1021 TGGCTGGGCTCTCTCAAGTCCAAATCGGCGACACCATGGCTGCGCGGCGTGTGGTGGG 1080
Db 3278 TGGCTGGGCTCTCTCAAGTCCAAATCGGCGACACCATGGCTGCGCGGCGTGTGGTGGG 3337
```

```
QY 1081 GTCATCAAGATGCTGATGGCTTCGCGGAGGGGGTGTGCCCGGACGTTGTCATGTGGAT 1140
Db 3338 GTCATCAAGATGCTGATGGCTTCGCGGAGGGGGTGTGCCCGGACGTTGTCATGTGGAT 3397
QY 1141 AAGCGCTCGCGCAGGTGCACTGGTCCGCGGGGGGGTCCGCTGCTGACGAGCGGGTG 1200
Db 3398 GAGCGCTCGCGCAGGTGCACTGGTCCGCGGGGGGGTCCGCTGCTGACGAGCGGGTG 3457
QY 1201 CCGTGGCGGGGACCGCGCAGGCGGCTTCGCGCGGGGGAGTGTCTGCTGTCGGGATC 1260
Db 3458 CCGTGGCGGGGACCGCGCAGGCGGCTTCGCGCGGGGGAGTGTCTGCTGTCGGGATC 3517
QY 1261 GCGCGCACGAATGCGCATGTGATTTTGAGGAGCGCGCGCG 1302
Db 3518 GCGCGCACGAATGCGCATGTGATTTTGAGGAGCGCGCGCG 3559

RESULT 4
AF275943
LOCUS
DEFINITION
  Streptomyces avermitilis avermectin polyketide synthase gene,
  partial cds.
ACCESSION
  AF275943
VERSION
  AF275943.1 GI:9964075
KEYWORDS
  Streptomyces avermitilis.
  Streptomyces avermitilis.
  Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
  1 (bases 1 to 11096)
  Hong, Y.-S. and Lee, J.J.
  Targeted Gene Disruption of the avermectin O-methyltransferase gene
  and polyketide synthase gene from Streptomyces avermitilis
  Unpublished
  2 (bases 1 to 11096)
  Hong, Y.-S. and Lee, J.J.
  Direct Submission
  Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
  Research Institute of Bioscience and Biotechnology, P.O. Box 116,
  Yuseong-Gu, Taejeon 305-600, South Korea
FEATURES
  Location/Qualifiers
    1..11096
      /organism="Streptomyces avermitilis"
      /strain="ATCC31271"
      /db_xref="ATCC:31271"
      /db_xref="taxon:33903"
      220..>11096
        /note="Avr1"
        /codon_start=1
        /transl_table=11
        /product="avermectin polyketide synthase"
        /protein_id="AAG09812.1"
        /db_xref="GI:9964076"
        /translation="MADEADGGVVVFPGQGPQWPGMGRELLDASDVRESVRACEAE
        PAPYVMSVEQVLRSDPADGLRDVVDVQRTLFAMISLAALMSVEPCVAGLHSL
        GEIAAAHVSGGLSLADAAARDVADWSAQTTLAGTALVSAATDELLPRTAPMTEDN
        PARLAAVANGPRSTVVGAREAVADLADLTAAQVTRMTDVDVPSHPLMLAIEER
        VVSGLLPITPRSPRPHSSVTGRLDRELDAAVWYRNMSSTVRFEPFAKLLLOQGP
        RTFVEMSPHPLVMTGLQELADLGLDVTGADTVIMGTLRGQGLDHLFTSLAQLRGH
        GFTSATVLSARLSTALSTQOSSLLDLVRATVWLVNDDGNRTASDAGSAGFAHL
        GFDSVMGVELRNLSKATGLRPVTLILDHTTAAVAARLTAALGHLEDDTAPVDS
        PSGHGTTAAADDPITAIIGMCRPPGVSRPKDLWELAAAGGDAIGPPPTDRGWPTQOR
        HAQDTPQGTFFYCGGGLHDAHFAGFFGISPREALAMPQORLLLETSTWEAPERA
        GIDPLSVRSRTGVFAGALSFDYGPMDTASGEAADVEGHLITGTGTVLSGSRAYS
        FGLGEPAITVDITGCSLVLTLHLACQSRSGECTFALAGGSDVDVHPGMFIBFSQCGI
        SVDCRCXAVSAADGTGCGEIGLLAVRGSVAVNDDGNRTASDAGSAGFAHL
        ANAGLSVADVVDVEHGTGTLGDIIEAQLLATYQORAGDRPLWGLSKNIHTMA
        AAGVGGVIMVMALREGVLPRTLHVDPSPQGLAAGVRLLTTEAVPWPDAAGRLRRA
        GVSSFGIGGTNAHVILEEAPAGCGVAGGVELEGAGLISVAESVAPVAPVAE
        SVVPVVPVVPVPSARSEAGLRAQAEALRQYEAQVDVSLDVAGLACRQAVLEHH
        VVILAACTSSSRSSAARTTARSSTSTARPOARPAPTSAKKI GALAAGSSGSAULTTGHAPG
        GDRGVVVFVFWQGGWAGMGRVLLCLLRVFAARMQACEEALAPFWMSVVDIILRRDA
        GDVAEQADVVQVPLVFSVMVSLAALWRSYGIENEVILGHSKDEITAAAHYGALSILKDA
```



```
/gene="aveA1"
101..12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module
2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="WQRMDCGEPRPAAGEVLGVADADGGVVFVPPGQGPQWGMGR
ELLDASDFRESRACEAFAPYVDMSVEQVLRSDPADGLDRVDVQPTLFAVMISL
AALMSQGVPECAVLGEIETAAHVSGLSLADAARVVTILWSQAQTTLAGTQALVS
VAATDELLPTAPTEDNPARLAAVANGPRSTTVSGAREAVADVADLTAAQVTRT
MIPDVPAPSLMYALIEBRVSGLLPITPRPSRIIPHSSVTGGRDTRELDAAWYRN
MSTVRFEPAAELLQOQPKTEVMSPEVLTMGLQELAPDLGDTGTGADTVMGTLR
RQGGILDHFLTSIAQLRGHGETSATIVLSARITALSPTQOQSLILLDLVRAHTMAVIND
DGNERTASDAGSPSAFHILGDFSVMGLVNRSLKATGLRPVTLIFORTTTPAAVAAR
LRTAALGHLEDPTAPVPSPSGGGTAAADPIAIIIGMCRFPGGVRSPKOLWEI LAAS
GDAIGPPTDRGWTPEQRHAQDTPQGTFFYQGGGFLHDAAHFDAGFFGISPRELAA
MPQOIRLLETTSWEAFERAGIDPLSVRSRTGVFFAGALSFDYPRMDTASSGGAADVE
GHILGTGSLSGRIAYSPGLEGPAITVDTCGSASLYTLHLACQSLSGECTLALAG
GVMSWTGLNFEFRQRLSDVGRKAYTAAADGTGMGEGVGLKVERLSDAVRIGH
RVLA VRGSAVNQDGSAGSLTAPNGPAQERVIRQALANAGLSVADVDDVVEGGTETTL
GDPIEAQALLATYQQRAGDRPLWGLSKSNI GHTMAAGVGVI KMVMALREGLVPT
LHVDXPSPOVDSAGAVRLLTEAVPMDGAAQLRAGVSSFGIGITNAHVILBEPAA
AGCVAGGVLSEAGPGLAISVAESNAAPVAVAPAESVPVPVPPVPPVARSSEAG
LRAQEAALRQYAVPVDVSLADVAGLACGRAVLEHRAVVLAAADREELVQGLGALAAG
EDPRRTIGHAGPDRGGVVFFPGQGGMGVRRLLASVFAVRMAQCEALAPW
VDMVVDLILRRDAGDAVNERADVQVPLFSVMVSLAALWRSYGIEDVILWQGEIJA
AAHVCGALSADAKTVALRSALAAVRGCGVASVLPQAQVEGLIGERWAGRLWA
AVNGPRSGADAEVLADEVLCAYCAGTGVRARRIIPDVASHCPHVPDAGVGLLELLE
ISQPSGVPEFSTVGTWLDITTLDAAYWYRNLHQPRESDAVQALADGHRVFEVS
PHTLVPALIEDTETAEDVTAIGSLRGDNDTRFPTALAHHTTGITGTPTWHHY
THHTRPHDHLDTLPTFPFQHYWLESSQPGAGSGAGSGAGSGRAGTAGGTA
EYESDFDVARADLDTVAITLAPVPSAGLTVFPALSAHRHQDQARINNTYQET
WKPLTLPTTHQPHQNTLTAIPEQTTHPHITNLNLNHHHTTPTPLNLNTHTPQH
LHHTLHTRQOQNTLTAITGLLSLLADETTPHHPHTGTGLNLNLTLTQTHTOH
PTPLWYATNATTHPNNDPLTHPTQAOQWGLARTLLEHPHTHAGIIDLPTTPHT
LQHLTQTLQPHQTLARTTGTTHRLTPTLPTTHQPTPTTHGTTLTGGTGAL
ATHLHLLTHTQTLHLLTSRTGHTTHAQHLTQLOQKQHLHLLTTCDTNSPDLQ
QUNLTIPPOHPTLVHTAGILDDATNLNLTPTQLNVRLAKASHALHQUTQHTPL
TAFVLSYSAATFTGAPQOYAAANAYDALAHRHTHLHPATSAMGTWQOGLADS
DKARAYLDRGRPRMPELATAVTOAIDTERPVVVIADIMSKIHTSOTSDLSVA
ABERPAVQPTPRAELHKTLAHOTSAQORALLERVDHVAALRHADPKAIAPDOQ
FRALGDSLTAVERFNLLIKATGLRPLVSLFDHTPTAKLAVHLQNLQRLGRAESAPS
AAVATABASVTBPIALVGNACRPPGGVTSADDFDLISSEQDAIGGFTDRGDWDLTL
YDPPDPHPTGTCYTRNGGLYDRHFDFAEFFGISPREALAMPQORLLLETAWETIEHA
GINPHTLHGTPTGVTGNGQDYALRVHNAQSTDGFALTGTAGSVISGRISYTFGE
GPAVDVTAACSSILVALHLACALRAGCSMALAGVTVMSPPGAVFVSFQRGLAAD
CHKAFSAADGTGEGVGLMLVERLSDAHNGHRVLAVRGSAVNODGASNGLTAP
NGPSQOVRITQALANAGLSAGVDVAEARGHTTGLDPIEAQALLATYQQRAGEGGL
WIGSVKSNVGHYTOAAAGVAGVIKMVMALRGLLPRTLHVDESPHVDMSAGVQLLTE
TVWPGCEGRLRAGVSSFGVSGTNAHVILEAPADDPVPGPAGEGAGDDEAAAG
SPGVNFWLVSQSPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDRHATLIA
ADHRTFLOALQAAGEHPHAVIHSSAPGGTGTGEAAKTAFCISCGQTOREPMAHGL
VHTHVEFAALNDICTHLDPLHDPLLPALTQNDNDNEDAALLQOTRYAPALFAFQ
VALHRLTDGYHITHTYAGSLGHEITAAHLAGILTLDTATTLTQRAILHLLHAGTY
TEHPLLGLTLELATDGGALLAGRLSIRSHPMIADHVGTVLISGATLELALHAGTY
VGCDDRDELTHAPVVDGVSQVGVAAADGGERLRLSVYARGGACGGGASGG
VMTCHASGVLEAAAGVVVDGLAGVPPRGAIVDVGDRDLRAGCGVLGVPVPSGL
RAVRMDGGDLLEVCLPEEAGDGAAGFLGHPALLDGVVQVLSVLLPGGTGCEGAFG
EGVRYPVAVMGVSLHRAGTVVRVRSVAVRGGRGEAVSVVVDGABGVASVDRLEL
RPVDMQGLRAVSAGRGSLYAVOWAEVGPVPCGQAWAHEDESGSGGVPVGVV
LRCPDAGAGGSGGGGGGAEVVGVLGVVQGLCLERFAGSRILVYTRGAVAGPE
DGFVDVGVWGLVRSQAQEDRFRVLLDLDITDGTDLDTGAGAGWGVGDRGAUV
ACGEPQLAVRGLAALKRLESSGDVPAQRSGTRARRSDVPAQRSGVPARRSVD
VSGREVLFWLSSGSLVLTGTTGVLGAARHLAGVCGVRDLLLLVSRRGPDAPGAEGLR
```

```
ABELAALGAEVRIVACDGERREVRVRLLEGVAGCPCLTGVVHAAGVLDATIASLTPER
LGTVPAAKVDAALLDELTRGMELSAFVLFSSAGCGQNTAANAALDALAYR
RRAGLPCVSLAWGLWEASGWTGHLACTHRRIRISGLHPMSTADALADALALD
RPVLLPADLRPAPPLPPLLOLLPATRRRTTTRTTTGGADNGAQLHLAGOTHQOH
TTLLALVESHIAVLGHNTPTDTPIDPRAFRDLGDSLTAVELNRLSRTTGLRLPTLL
AFDHPNPTLLTHLHTQLOPQDNAPVAFVLAELDKLSALDKTDSASERTVLRKL
SLMLRWNAPOHPTAESADDBKFTSATEABEIKFIDNDLGLS"
12071..30790
/gene="aveA2"
12071..30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and
module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="WQLANEAKLLEYLKRVTADLDRTRRLRYEVVEREQEPIAVGMA
CRYPGCAFTRLWHLVKSQDAIGEFPTDRGMNLQOLYDPPDRSGTSTRSGGFLY
DAGFDAAFPFELSPREALAMPQORLLLETTWEITEQGIIDPRSMRGSRGTGVGINP
CRYPGCAFTRLWHLVKSQDAIGEFPTDRGMNLQOLYDPPDRSGTSTRSGGFLY
EDYTTGYTHQPSNAVEGYLLTGSASAISAGRISYNFLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVWATPVFTFQRQGLAAGRCATFSAADGTGSGEGV
ACMLVERLSDARRNGHRVLAVRGSAVNQDGSAGSLTAPNGRSQVYKIRQALANHL
PADVDVAEHGTTGLDPIEQAQALVEAYGQDRPNRPLWGLTGLTNSKNGHMAAGVG
GVKVMALNRGLLPTLHVBPSPGHVDSAGAVOLLTETVMPGCEGRLLRAGVSSF
GVSTNAHVILEEPAHNI PSDTPADDDVPGESADEDEDAGSGDEAAGSPVWPLVSA
KSQAPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDRHATLIAADROTFFLOALQ
ALAAEPHPAVIHSSAPGGTGEAAKTAFCISGQTOREPMAHGLVHTHVPFAAAL
NDICTHLDPLHDPLLPALTQDPTNTDTTLEEAALLQOTRYAPALFAFQVALLHRL
LTDGVIPTPHYAGSLGEITAAHLAGILTLDTATTLTQRAILHMQTMDPGTMTLHT
TPHITLHHTAHENDLATAINTPTSLVISGTPHVOHITTLCOQKQIKTKLTPNHA
FHSPTNPIANQLHQHTQTLVHPHPTPLTANTPPDQLLTPHVTQOABNTVDYATT
TQTLHQGVITYIELGPDNTLTLTHDMLNPTPTTLTLTHPHHPQHLNLNLAKT
TWPHHYTHHNPHTHLLDPTYPFQHHHYWLPQPKGSPSGSDESGEQATTPST
PLRDLVKGSPQERDEELLRLVTHAAVLAHATPEVVPNKAFKELGDSLAAIQLR
NMLADVDLPLPATLIDFYPPTMALCQFLRAAIVGADTGTTRTLPPLTAPADEPIAV
GMACRYPGCVTRVDLMOVVSQGHDAIGGPTNRQMDLDTLYNPPDRHHGTYSTRSG
FLYDAGNDFDPFGISPREALAMPQORLLLETAWETIEHACINPDSIRGTPGVPA
LTYDHYAARFTAPAGFEGYLGHSAGSIASGRVAYALGLEGPALTVDITACSSSLVAL
HLACQALRSGECMALAGGVTVMSTPAGFVSFQRQGLAAGRCATFSAADGTGSGEG
GVKVMALNRGLLPTLHVBPSPGHVDSAGAVOLLTETVMPGCEGRLLRAGVSSF
LTPADVDVAEHGTTGLDPIEQAQALVEAYGQDRPNRPLWGLTGLTNSKNGHMAAG
VGVYKVMALNRGLLPTLHVBPSPGHVDSAGAVOLLTETVMPGCEGRLLRAGVSS
SVGVTGNAHVILEEPAHNI PSDTPADDDVPGESAADDDAGSGDEAGTGEAGTGA
AGSGVWPLVSAQSPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDRHATL
IAADROTFFLOALQAAGEHPHAVIHSSAPGGTGEAAKTAFCISGQTOREPMAH
GLYTHVPFAAALNDICTHLDPLHDPLLPALTQDPTNTDTTLEEAALLQOTRYAQ
PALFAFQVALLHRLTDGVIPTPHYAGSLGEITAAHLAGILTLDTATTLTQRAILM
QTPMGVTWTLTHTPHITLHHTAHENDLATAINTPTSLVISGTPHVOHITTLCOQ
QGIKTKLPTKNAFHSPTNPIANQLHQHTQTLVHPHPTPLTANTPPDQLLTPHYM
TQOABNTVDYATTQTLHQGVITYIELGPDNTLTLTHDMLNPTPTTLTLTHPHHP
PQTHLLNLAKTTTTHPHHYTHHNPHTHLLDPTYPFQHHHYWLPQPKGSPSGSDE
SGSGSGRAGTAGTAEVSRFWDVARDQLETVATLAVPPSAGLTVPLASAWHRH
QHDQARINNTYQETWKPLTPTTHQHTWLIAPETQTHPHITNLNLNHHHTT
PIPLTANHTNPOHLHTRQOQNTLTAITGLLSLLADETTPHHPHTGTGLNLNL
LTLTQTHPTPLWYATNATTHPNNDPLTHPTQAOQWGLARTLLEHPHTHAGI
IDLPTTPHTLHLLTQTLQPHQHTQTLARTTGTTHRLTPTLPTTHQPTPTTHG
TTLTGTGALATLHLLTHTHPTQHLHLLTSRTGHTTHAQHLTQLOQKQHLHLLT
TCDTNSPDLQUNLTIPPOHPTLVHTAGILDDATNLNLTPTQLNVRLAKASHAL
LLHQUTQHTQTLHPTVLSAAATFTGAPQOYAAANAYDALAHRHTHLHPATSIAW
GTWQOGLADTQVSEHLRRCMGFPPELAVTAVDGAIASGRPSLLVADLWPKGLPV
LSSKSVLLELPOAQGTGEARSTVEQTESNLQRLMGRSRSQEEBELLISVIRHSA
AVLGRDSEAIIPGRLFRDUGFDSLAEVLAHVLAHQTELALPTLVLFYDPSFTPLAQ
FLLSIEAIEFQDNSTPLPRPAELDEPIAVIAGMCRFPGGVTSADDFDLISSEQDA
GGPTDRGMDLTDYDPPDPHPTGTCYTRNGGLYDRHFDFAEFFGISPREALAMPQ
RLLETAWETIEHAGINPHTLHHTQHTGTPVGTCTNGQDHAAHIRQAQPSGTFVLTCAAT
SIASGRISYILGLESGPALTATCSSILVALHLACQSLSGECTMALAGGATVMTPI
TTFTEFQRGLAPDGRCKASAAADGTGAGEGVGLMLVERLSDARRNGHRVLAVRGS
AVNODGASNGLTAPNGPAQERVIRQALANAGLSVADVDAEHGTTGLDPIEQAQ
LATYQQRPGNGLAALKRLESSGDVPAQRSGTRARRSDVPAQRSGVPARRSVD
HVDMSAGAVOLLTETVMPGCEGRPRRAGVSSFGVSGTNAHVILEEAPADDPVPGGPPA
```

THE KITASATO INSTITUTE	
OS	Streptomyces avermitilis
PN	JP 200245457-A/1
PD	12-SEP-2000
PF	24-FEB-1999 JP 1999046961
PI	SATOSHI OMURA, HARUO IKEDA
PC	C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
PC	C12Q1/68/C07D493/22
PC	(C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
Key	Location/Qualifiers
FT	CDS (11916)
FT	CDS (11971), (30687)
Location/Qualifiers	
1..30690	
/organism="Streptomyces avermitilis"	
/db_xref="taxon:33903"	
BASE COUNT	5355 a 12455 c 8617 g 4263 t
ORIGIN	
Query Match 100.0%; Score 1302; DB 6; Length 30690;	
Best Local Similarity 100.0%; Pred. No. 2.9e-155;	
Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 GACCGATCCCATCATCGCATGCGATGCGTTCCTCCCGGCGGAGTCCGGTCCCGAAG 60
Db	1441 GACCGATCGCCATCATCGCATGCGATGCGTTCCTCCCGGCGGAGTCCGGTCCCGAAG 1500
Qy	61 GACTGTGGAGTGGCCGCTCGGCGGAGAGCCATCGGGCGTTCCTCCCAACGACCGC 120
Db	1501 GACTGTGGAGTGGCCGCTCGGCGGAGAGCCATCGGGCGTTCCTCCCAACGACCGC 1560
Qy	121 GATGCCCCACAGAGCGTCACGCCAGAGCCACCCAGCCAGCGTTCATCCG 180
Db	1561 GATGCCCCACAGAGCGTCACGCCAGAGCCACCCAGCGTTCATCCG 1620
Qy	181 CAGGAGGCGGTTCTTCCACAGCGGCGCACTTCGACCGCGGCTTCTTCGGAATCACT 240
Db	1621 CAGGAGGCGGTTCTTCCACAGCGGCGCACTTCGACCGCGGCTTCTTCGGAATCACT 1680
Qy	241 CCACGTGAGCACTGGCGATGATCCGACGAGCGGCTGCTGTGGAGAGCTCTGGAG 300
Db	1681 CCACGTGAGCACTGGCGATGATCCGACGAGCGGCTGCTGTGGAGAGCTCTGGAG 1740
Qy	301 GCGTTCAGCGGCGGGAATCGATCGCTGTGGTACGCGGCTCCGTACGCGGCTTC 360
Db	1741 GCGTTCAGCGGCGGGAATCGATCGCTGTGGTACGCGGCTCCGTACGCGGCTTC 1800
Qy	361 GCGGCGCCCTCTCTTCGACTACGCGCCCGGATGACACCGCGTCTCGGAGGCGCC 420
Db	1801 GCGGCGCCCTCTCTTCGACTACGCGCCCGGATGACACCGCGTCTCGGAGGCGCC 1860
Qy	421 GCGGAGTGGAGGCGCACATCTCACCGGTACACGGGAGCGTCTGTGGGCGGTATC 480
Db	1861 GCGGAGTGGAGGCGCACATCTCACCGGTACACGGGAGCGTCTGTGGGCGGTATC 1920
Qy	481 GCTACAGTTCGGGCTGGAAGGCGCGGATCACCGTGGACACGCGGCTCGGCATCG 540
Db	1921 GCTACAGTTCGGGCTGGAAGGCGCGGATCACCGTGGACACGCGGCTCGGCATCG 1980
Qy	541 CTGCTGACGCTGATCTGGCGTCCAGTCTGCTCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db	1981 CTGCTGACGCTGATCTGGCGTCCAGTCTGCTCGGTCGGGTGAGTGCACGCTCGCGCTG 2040
Qy	601 GCGGCGGCTCTCGGTATGTCACCTCGGATGTTATCGAGTTCCTCCGCGACCGC 660
Db	2041 GCGGCGGCTCTCGGTATGTCACCTCGGATGTTATCGAGTTCCTCCGCGACCGC 2100
Qy	661 GGGCTCTGCTGGACGCGAGTGCAGGGGCTACTCGGCTGCAGCCGACCGCGCTCG 720
Db	2101 GGGCTCTGCTGGACGCGAGTGCAGGGGCTACTCGGCTGCAGCCGACCGCGCTCG 2160
Qy	721 GCGAGGCGCTCGGATGCTGTTGGTGGAGCGGTTGTCGATGCGGCTGGGGCAT 780
Db	2161 GCGAGGCGCTCGGATGCTGTTGGTGGAGCGGTTGTCGATGCGGCTGGGGCAT 2220
Qy	781 CCGGTGCTGGCGGTGTACCGCGAGTCCGCTCAACAGGACGCTGCTCAATGGGCTG 840
Db	2221 CCGGTGCTGGCGGTGTACCGCGAGTCCGCTCAACAGGACGCTGCTCAATGGGCTG 2280
Qy	841 ACGGCCCGAAACGGTCCCGCTCAGAGCGGCTGATCCCGCAGGCGTTGGCGAAACGCGGG 900
Db	2281 ACGGCCCGAAACGGTCCCGCTCAGAGCGGCTGATCCCGCAGGCGTTGGCGAAACGCGGG 2340
Qy	901 TTGTCGCTGGCGGATGTGGATGTGGTGGAGGGGACCGGACCGGACGCTGGGTGAT 960
Db	2341 TTGTCGCTGGCGGATGTGGATGTGGTGGAGGGGACCGGACCGGACGCTGGGTGAT 2400
Qy	961 CCGATCGAGGACAGGCGTTCCTCCACGCTACGGGACGCGCGGCTGACAGGCGGCTG 1020
Db	2401 CCGATCGAGGACAGGCGTTCCTCCACGCTACGGGACGCGCGGCTGACAGGCGGCTG 2460
Qy	1021 TGGCTGGGCTCTCTGAAGTCCAAACATCGGCGACACCATGGCTGCCGCGGCTGTGGGTGG 1080
Db	2461 TGGCTGGGCTCTCTGAAGTCCAAACATCGGCGACACCATGGCTGCCGCGGCTGTGGGTGG 2520
Qy	1081 GTCATCAAGATGTGATGCGCTTGGGAGGGGTTTCCCGCGGACGTTGCATGTGGAT 1140
Db	2521 GTCATCAAGATGTGATGCGCTTGGGAGGGGTTTCCCGCGGACGTTGCATGTGGAT 2580
Qy	1141 AAGCGCTCGCGCAGGTGACTGCTCCGCGGGGGGTCGCGCTCTCACGAGGCGGTTG 1200
Db	2581 AAGCGCTCGCGCAGGTGACTGCTCCGCGGGGGGTCGCGCTCTCACGAGGCGGTTG 2640
Qy	1201 CCGTGGCGGGGACCGCGCAGGCGGTTGCGCGGGGGGTCGCGCTCTCACGAGGCGGTTG 1260
Db	2641 CCGTGGCGGGGACCGCGCAGGCGGTTGCGCGGGGGGTCGCGCTCTCACGAGGCGGTTG 2700
Qy	1261 GCGCGACGAATGCGCATGTGATTTTGAGAGGCGCGCGG 1302
Db	2701 GCGCGACGAATGCGCATGTGATTTTGAGAGGCGCGCGG 2742
RESULT 2	
AB032367	64957 bp DNA linear BCT 18-SEP-1999
LOCUS	Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
DEFINITION	AB032367.1 GI:5902890
ACCESSION	AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES 4; type I polyketide synthase AVES 3; type I polyketide synthase AVES 2; type I polyketide synthase AVES 1.
VERSION	Streptomyces avermitilis DNA.
KEYWORDS	Streptomyces avermitilis
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
ORGANISM	1 (bases 1 to 64957)
REFERENCE	2 (bases 1 to 64957)
AUTHORS	Ikeda,H., Nonomura,T., Usami,M., Ohta,T. and Omura,S.
TITLE	Organization of the biosynthetic gene cluster for the polyketide anthelmintic macroide avermectin in Streptomyces avermitilis
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
MEDLINE	99380548
REFERENCE	2 (bases 1 to 64957)
AUTHORS	Ikeda,H., Nonomura,T., Usami,M., Ohta,T. and Omura,S.
TITLE	Direct Submission
JOURNAL	Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical Sciences, Kitasato University, Microbial chemistry; 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
FEATURES	(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)
source	Location/Qualifiers
gene	1..64957
	/organism="Streptomyces avermitilis"
	/db_xref="taxon:33903"
	101..12019

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 2525.62 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742

Perfect score: 1302
Sequence: 1 gaccgatgcgcattcatcg...ttttggaggagcgccgcggc 1302

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		DB	ID	Description	
		Match	Length				
1	1302	100.0	30690	6	E38020	E38020 Avermectin	
2	1302	100.0	64957	6	AB032367	AB032367 Streptomy	
3	1298.8	99.8	12381	6	AX006889	AX006889 Sequence	
4	1128.8	86.7	11096	1	AF275943	AF275943 Streptomy	
5	944.8	72.6	31422	6	E38021	E38021 Avermectin	
C	6	944.8	72.6	64957	1	AB032367	AB032367 Streptomy
	7	738.6	56.7	78210	1	AB070949	AB070949 Streptomy
8	722.2	55.5	41782	1	AY118081	AY118081 Streptomy	
9	710	54.5	20394	1	SNA13222	AJ132222 Streptomy	
C	10	710	54.5	20394	6	AX067996	AX067996 Sequence
	11	710	54.5	84985	1	SNA278573	AJ278573 Streptomy
12	689	52.9	24568	1	AF220951	AF220951 Streptomy	
13	689	52.9	50937	6	AR159871	AR159871 Sequence	
C	14	688	52.8	104326	1	AB070940	AB070940 Streptomy
	15	677	52.0	30000	6	AX250262	AX250262 Sequence
16	676.4	52.0	43280	1	SFU78289	U78289 Streptomyce	
17	675.8	51.9	104326	1	AB070940	AB070940 Streptomy	
18	674.6	51.8	25150	1	AB070945	AB070945 Streptomy	
19	672.6	51.7	30000	6	AX250263	AX250263 Sequence	
20	662.6	50.9	5544	1	SNA13221	AJ132221 Streptomy	
21	662.6	50.9	5544	6	AX067994	AX067994 Sequence	
22	662.6	50.9	84985	1	SNA278573	AJ278573 Streptomy	
23	659.8	50.7	11480	1	STMPK3ORF	L09654 Streptomyce	
24	655	50.3	53784	1	AMW223012	AJ223012 Amycolato	
25	655	50.3	53789	6	A69720	A69720 Sequence 3	
26	655	50.3	90445	1	AF040570	AF040570 Amycolato	
27	652.8	50.1	3900	1	AF262754	AF262754 Amycolato	
28	650.4	50.0	14775	6	AX089462	AX089462 Sequence	
29	650.4	50.0	50000	6	AX089417	AX089417 Sequence	
30	650.4	50.0	50000	6	AX089420	AX089420 Sequence	
31	650.4	50.0	80161	1	AY007564	AY007564 Saccharop	
32	650.4	50.0	80161	6	AR165018	AR165018 Sequence	
33	649	49.8	16767	6	AX089464	AX089464 Sequence	
34	646.4	49.6	9513	6	AX089460	AX089460 Sequence	
C	35	645.8	49.6	24225	1	SC2C4	AL512902 Streptomy
	36	645.4	49.6	6459	6	AX089458	AX089458 Sequence
C	37	641.4	49.3	82746	1	AF453501	AF453501 Actinosyn
	38	639	49.1	37948	1	AF079138	AF079138 Streptomy
39	638.2	49.0	27541	6	AX211706	AX211706 Sequence	
40	638.2	49.0	123580	1	AF263912	AF263912 Streptomy	
41	638.2	49.0	125401	6	AX211739	AX211739 Sequence	
C	42	637.4	49.0	107379	1	SHGCP1R	X86780 S.hygrosco
	43	636.4	48.9	41097	1	AF016585	AF016585 Streptomy
C	44	636.2	48.9	26195	1	SC1G7	AL591083 Streptomy
	45	636	48.8	65140	6	AX211705	AX211705 Sequence

ALIGNMENTS

RESULT 1

E38020	E38020	30690 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	Avermectin aglycon synthase gene.				
DEFINITION	E38020				
ACCESSION	E38020				
VERSION	E38020.1	GI:18626909			
KEYWORDS	JP 2000245457-A/1.				
SOURCE	Streptomyces avermitilis.				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
REFERENCE	Streptomycineae; Streptomycetaceae; Streptomycetes.				
AUTHORS	1 (bases 1 to 30690)				
TITLE	Omura,S. and Ikeda,H.				
JOURNAL	Avermectin aglycon synthase gene				
	Patent: JP 2000245457-A 1 12-SEP-2000;				

THIS PAGE BLANK (USPTO)

Qy 245 TC 246
Db 448 CC 447

Search completed: June 18, 2003, 01:00:34
Job time : 318.438 secs


```

QY 5 CCGGGCAGACACGAAACAGACACACACCTCTCGCCCTCGTCCGTCCTCCACATCG 64
Db 834 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 775

QY 65 CCACCGTCTGGGCGCACACACCCCGGACACCATCCCGCCGCGGTTCCGGGACC 124
Db 774 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 715

QY 125 TCGGCTTCGACTCCCTCACCGCGTGAACCTAGCAACCGGCTCTCCCGCACACCGGAC 184
Db 714 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 655

QY 185 TCGGCTCCCGACACCTCGCTTCGACACACCCCAACCCACACCTCACCCACAC 244
Db 654 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 595

QY 245 TC 246
Db 594 CC 593

RESULT 8
CNS04FIQ
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
106L07 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL288395
VERSION
GI:8026928
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 389)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL
2 (bases 1 to 389)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL
3 (bases 1 to 389)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end and sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..389
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="106L07"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGI06CF04SP1-end :
PUC-Ori"
BASE COUNT 34 a 279 c 34 g 29 t 13 others
ORIGIN
Query Match 32.2%; Score 79.2; DB 17; Length 389;
Best Local Similarity 55.0%; Pred. No. 5.8e-05;
Matches 132; Conservative 10; Mismatches 98; Indels 0; Gaps 0;

QY 5 CCGGGCAGACACGAAACAGACACACACCTCTCGCCCTCGTCCGTCCTCCACATCG 64
Db 834 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 775

QY 65 CCACCGTCTGGGCGCACACACCCCGGACACCATCCCGCCGCGGTTCCGGGACC 124
Db 774 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 715

QY 125 TCGGCTTCGACTCCCTCACCGCGTGAACCTAGCAACCGGCTCTCCCGCACACCGGAC 184
Db 714 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 655

QY 185 TCGGCTCCCGACACCTCGCTTCGACACACCCCAACCCACACCTCACCCACAC 244
Db 654 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 595

QY 245 TC 246
Db 594 CC 593

RESULT 8
CNS04FIQ
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
106L07 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL288395
VERSION
GI:8026928
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE
1 (bases 1 to 389)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL
2 (bases 1 to 389)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
JOURNAL
3 (bases 1 to 389)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end and sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..389
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="106L07"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGI06CF04SP1-end :
PUC-Ori"
BASE COUNT 34 a 279 c 34 g 29 t 13 others
ORIGIN
Query Match 32.2%; Score 79.2; DB 17; Length 389;
Best Local Similarity 55.0%; Pred. No. 5.8e-05;
Matches 132; Conservative 10; Mismatches 98; Indels 0; Gaps 0;

```

```

QY 7 GCGGAGACACAGAAACAGACACACCTCTCGCCCTCGTCCGTCCTCCACATCGCC 66
Db 117 GGTTCACACACACACACACACACACACACACACACACACACACACACACACAC 176

QY 67 ACGGTCCTGGGCGCACACACCCCGGACACCATCCCGCCGCGGTTCCGGGACCTC 126
Db 177 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 236

QY 127 GCGTTCGACTCCCTCACCGCGTGAACCTAGCAACCGGCTCTCCCGCACACCGGACTC 186
Db 237 CCCCCCCCCCCCCCTCYCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 296

QY 187 CCGCTCCCGACACCTCGCTTCGACACACCCCAACCCACACCTCACCCACACCTC 246
Db 297 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 356

RESULT 9
BG786282/c
LOCUS
DEFINITION
SEAUMC006239 Sea urchin primary mesenchyme cell cDNA library
Strongylocentrotus purpuratus cDNA clone PC_0028_A2_C11_MR 5', mRNA
sequence.
ACCESSION
BG786282
VERSION
BG786282.1
KEYWORDS
EST.
SOURCE
Strongylocentrotus purpuratus.
ORGANISM
Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinoidae; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE
1 (bases 1 to 646)
Zhu, X., Mahairas, G., Illies, M.R., Cameron, R.A., Davidson, E.H. and
Ettensohn, C.A.
A large scale analysis of mRNAs expressed by primary mesenchyme
cells of the sea urchin embryo
Development 128 (13), 2615-2627 (2001)
JOURNAL
MEDLINE
21384984
COMMENT
Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
FEATURES
Location/Qualifiers
1..646
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC_0028_A2_C11_MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA
library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/notes="Vector: pSPori1; Site 1: NotI; Site 2: SalI; oligo
dt priming from poly A+ RNA, directionally cloned"
BASE COUNT 2 a 3 c 566 g 50 t 25 others
ORIGIN
Query Match 32.1%; Score 79; DB 12; Length 646;
Best Local Similarity 56.5%; Pred. No. 5.9e-05;
Matches 139; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 1 CTGGCGCGCCAGACACAGCAACAGCAGACACACCTCTCGCCCTCGTCCGTCCTCCAC 60
Db 545 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 486

QY 61 ATCGCCACCGTCTTGGGCGCACACACCCCGGACACATCCCGCCGACCGCGGTTCCGC 120
Db 485 CCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 426

QY 121 GACCTCGGCTTCGACTCCCTCAACCGCGCTCGAATACGCAACCGGCTCTCCCGCACCA 180

```



```

/strain="PEST"
/db_xref="taxon:7165"
/clone="22N11"
/clone_lib="NotreDamel"
/note="end : T7"
BASE COUNT      136 a      552 c      71 g      112 t      83 others
ORIGIN
Query Match      32.4%; Score 79.6; DB 17; Length 954;
Best Local Similarity 56.2%; Pred. No. 4.4e-05;
Matches 136; Conservative 6; Mismatches 100; Indels 0; Gaps 0;
QY 5 CCGGCCAGACACAGCAACAGCAGCACACACCTCTCGCCCTCTGCTGCTCCACATCG 64
DB 175 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 234
QY 65 CACACGCTCTGGGCGACACACACACACACACACACACACACACACACACACACAC 124
DB 235 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 294
QY 125 TCGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 184
DB 295 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 354
QY 185 TCGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 244
DB 355 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 414
QY 245 TC 246
DB 415 CC 416

RESULT 6
BI416474/c
LOCUS
DEFINITION
Asiegbu.F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 803)
AUTHORS
Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLES
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL
Unpublished (2001)
COMMENT
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
Location/Qualifiers
1..803
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xa09f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."
BASE COUNT      45 a      17 c      723 g      13 t      5 others

```

```

ORIGIN
Query Match      32.3%; Score 79.4; DB 13; Length 803;
Best Local Similarity 57.9%; Pred. No. 4.9e-05;
Matches 140; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 5 CCGGCCAGACACAGCAACAGCAGCACACCTCTCTCGCCCTCTGCTGCTCCACATCG 64
DB 668 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 609
QY 65 CACACGCTCTGGGCGACACACACACACACACACACACACACACACACACACACAC 124
DB 608 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 549
QY 125 TCGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 184
DB 548 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 489
QY 185 TCGGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 244
DB 488 CCGGCCAGACACACACACACACACACACACACACACACACACACACACACACAC 429
QY 245 TC 246
DB 428 TC 427

RESULT 7
BI416575/c
LOCUS
DEFINITION
Asiegbu.F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 931)
AUTHORS
Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLES
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL
Unpublished (2001)
COMMENT
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
Location/Qualifiers
1..931
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg15f"
/clone_lib="Heterobasidion annosum - Scots pine infection
stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FP5)."
BASE COUNT      9 a      7 c      884 g      15 t      16 others
ORIGIN
Query Match      32.3%; Score 79.4; DB 13; Length 931;
Best Local Similarity 57.9%; Pred. No. 4.8e-05;
Matches 140; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 315.347 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658

Perfect score: 246

Sequence: 1 ctggcgccgacacacga.....ccaccctaccaccacctc 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hct:*
9: gb_est1:*
10: gb_est2:*
11: gb_hct:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	85.2	34.6	825	12	BG786255
C 2	81.6	33.2	855	17	AG047947
C 3	80.6	32.8	688	12	BG786340
C 4	79.8	32.4	1228	12	BG067997
C 5	79.6	32.4	954	17	CNS01MZJ
C 6	79.4	32.3	803	13	BI416474
					BI416474 hasp001xa

C 7	79.4	32.3	931	13	BI416575
C 8	79.2	32.2	389	17	CNS04FIQ
C 9	79	32.1	646	12	BG786282
C 10	79	32.1	744	17	AG153752
C 11	79	32.1	912	17	AG136783
C 12	78.8	32.0	375	12	BG519974
C 13	78.8	32.0	585	10	AW335294
C 14	78.8	32.0	848	17	CNS0118A
C 15	78.8	32.0	913	17	AZ687957
C 16	78.6	32.0	806	13	BI416659
C 17	78.6	32.0	1087	17	AG043580
C 18	78.4	31.9	1101	17	CNS0181N
C 19	78.2	31.8	1303	14	BM908878
C 20	78	31.7	715	9	AL570062
C 21	77.8	31.6	738	17	AZ199465
C 22	77.8	31.6	818	17	AG060534
C 23	77.8	31.6	960	17	CNS0048W
C 24	77.6	31.5	586	17	CNS01UD1
C 25	77.6	31.5	669	17	CNS032YM
C 26	77.6	31.5	736	17	AZ197707
C 27	77.6	31.5	929	17	AZ687353
C 28	77.4	31.5	261	17	CNS04KHD
C 29	77.4	31.5	637	17	AG077880
C 30	77.4	31.5	823	17	AZ187132
C 31	77.4	31.5	1088	13	BM547778
C 32	77.4	31.5	1166	12	BF256751
C 33	77.2	31.4	529	10	AW050654
C 34	77.2	31.4	559	17	CNS04CPE
C 35	77.2	31.4	593	10	AW320968
C 36	77.2	31.4	638	17	CNS046XY
C 37	77.2	31.4	740	17	AZ195718
C 38	77.2	31.4	812	9	AL537183
C 39	77.2	31.4	843	17	AZ195427
C 40	77.2	31.4	903	17	BH136778
C 41	77.2	31.4	942	9	AL536132
C 42	77.2	31.4	945	17	AG117065
C 43	77	31.3	424	10	AW778904
C 44	77	31.3	459	9	A1285778
C 45	77	31.3	928	9	AL561014

ALIGNMENTS

RESULT 1
BG786255/c 825 bp mRNA linear EST 20-MAY-2001
LOCUS SEAMC006212 Sea urchin primary mesenchyme cell cDNA library
DEFINITION Strongylocentrotus purpuratus cDNA clone PC_0028_A2_A08_MR 5', mRNA sequence.
ACCESSION BG786255
VERSION BG786255.1 GI:14157268
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 825)
AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and Ettensohn,C.A.
TITLE A large scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo
JOURNAL Development 128 (13), 2615-2627 (2001)
MEDLINE 21384984
COMMENT Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
Location/Qualifiers
FEATURES
source 1. .825

Db	29173	CCCTGCGCGACCTCGTCCGGGGCCA	CGTCCGCTCGTGTCTCGGACACAGTGGCCCCCGAGG	29232
Qy	95	CCATCCCCCCGACCGCGCTTCCGGAG	CACTCGGCTTCGACTCCCTCACGCGCTCGAAC	154
Db	29233	CCATCGACGCGACACAGGCCTTCCGGGAC	CATCGGTTTCGACTCCCTGACCGCAGTTCGAAC	29292
Qy	155	TACGCAACCGGCTCTCCCGCACACCGAGACTCGGCTCCGCCACCACTTCGCTTCGACC	214	
Db	29293	TCAGAAACCGGCTGAACGCGGAGACCGGCTTCGGCTTCGCCGGCAGCTCGTGGTTCGACT	29352	
Qy	215	ACCCCAACCCACACCGCTCACCCACCACT	245	
Db	29353	ACGCCAACCCGAGCGGCTCGCGCATCACT	29383	

```

RESULT 11
US-09-860-846-34
; Sequence 34, Application US/09860846
; Patent No. US2002016742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-34

```

Query Match	45.2%	Score 111.2;	DB 9;	Length 4689;
Best Local Similarity	70.3%;	Pred. No. 1.1e-18;		
Matches 149;	Conservative 0;	Mismatches 63;	Indels 0;	Gaps 0;
Qy	35	CCCTCTCCGGCTGGTCCCGTCCACATCGCACCGTCTCTGGGGCCACACACCCCGGACA	94	
Db	4211	CGTCTCTACCGCTCGTCCGTACCCACCGGGGGCGCTACTTCGGCCATTCTCCCCGACC		4270
Qy	95	CCATCCCCCGGACCGCGCGTTCGGGACTTCGGCTTCGACTCCCTCACCGCGGTGGAAC	154	
Db	4271	GGGTGGGCCCGGGCGTGCCTTCACCGAGCTTCGGCTTCGACTCGTCAGCGCGGTGCAGC		4330
Qy	155	TAGCGAAACGGCTCTCCCGGACACACCGGACTCGCGCTCCCGACACACCTCGCTTCGACC	214	
Db	4331	TCGGCAACAGACTCTCACGGTGGTGGCAACAGGCTCCCGGCCACCGGTTCTTCGACC		4390
Qy	215	ACCCCAACCCGACACCTCAACCACCACTC	246	
Db	4391	ACCCGACGCGCGCGGCACTGGCGGCACTC	4422	

```

RESULT 12
US-09-988-384B-34
; Sequence 34, Application US/09988384B
; Publication NO. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398

```

```

; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 34.
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-34

```

Query Match	45.2%	Score 111.2	DB 9	Length 4689
Best Local Similarity	70.3%	Pred. No. 1.1e-18		
Matches 149	Conservative	0	Mismatches 63	Indels 0
Gaps	0			
Qy	35	CCCTCTCTCGCGCTTCGGTCCGTCACATCGCCACCGTCTCTGGGCCACACACCCCGACA	94	
Db	4211	CGTCTCTCACCTCTGTCGGTACCCACCGGGGCGTACTCGGCATTCTCTCCCGACC		
Qy	95	CCATTCGCCCCCGACCGGGGTTTCGGGACCTCGGCTTCGACTCCCTACCGCGGTGGAAC	154	
Db	4271	GGGTGGCCCCGGCCGTCGCTTCACCGAGCTCGGCTTCGACTCGTGCAGGGCGTGCAGC	4330	
Qy	155	TAGCAACCGGCTCTCCGGCACACCGGACTCGGCTCCCCACACCTTCGCTTCGACC	214	
Db	4331	TCGCGCAACAGCTCTTCACGGTGGTCGGACAGGGTCTCCGGCCACACGGTCTTCGACC	4390	
Qy	215	ACCCCAACCCACCAACCTTCAACCAACCTC	246	
Db	4391	ACCGGACGGCCGCGGACTCGCCGGCAGCTC	4422	

```

RESULT 13
US-09-836-821-34
; Sequence 34, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikomycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836.821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-34

```

Query Match	45.2%;	Score 111.2;	DB 9;	Length 4689;
Best Local Similarity	70.3%;	Pred. No. 1.1e-18;		
Matches 149;	Conservative 0;	Mismatches 63;	Indels 0;	Gaps 0;
Qy	35	CCCTCTCGCCCTGGTCGCTCCACATCGCCACCGTCTCTGGGCACACACCCCGACA	94	
Db	4211	CGTCTCTCACCCTCTGTCGTAACCGCGCGCGGTACTTCGGCCATTCTCCCCGACC	4270	
Qy	95	CCATCCCCCCCGACCGCGGTTCCGCGACTCTCGGTTGCATCTCCCTACCGCGCTGGAAC	154	
Db	4271	GGGTGGCCCCCGCGCGTGCCTTACCGAGCTCGGCTTCGACTCGCTAGCGCGCTGCAGC	4330	
Qy	155	TACGCAACCGGCTCTCCCGCACCACGGACTCGGCTTCCGACCAACCCTCGCTTCGACC	214	
Db	4331	TCGCAACACAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTTCGACC	4390	
Qy	215	ACCCCAACCCACCAACCCCTCAACCCACCACTC	246	
Db	4391	ACCCGACGCGCGCGGACTTCGGCGGCACCTC	4422	

QY		241	CACCT 245	
DB		10918	CACCT 10922	
RESULT 5				
US-09-860-846-5				
; Sequence 5, Application US/09860846				
; Patent No. US20020164742A1				
; GENERAL INFORMATION:				
; APPLICANT: Sherman, D.H.				
; APPLICANT: Liu, H.				
; APPLICANT: Xue, Y.				
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin				
; FILE REFERENCE: 600.438U51				
; CURRENT APPLICATION NUMBER: US/09/860.846				
; CURRENT FILING DATE: 2001-05-18				
; PRIOR APPLICATION NUMBER: 09/105.537				
; PRIOR FILING DATE: 1998-06-26				
; NUMBER OF SEQ ID NOS: 43				
; SOFTWARE: FastSeq for Windows Version 3.0				
; SEQ ID NO 5				
; LENGTH: 36778				
; TYPE: DNA				
; ORGANISM: Streptomyces venezuelae				
US-09-860-846-5				
Query Match 48.2%; Score 118.6; DB 9; Length 36778;				
Best Local Similarity 67.8%; Pred. No. 1.4e-20;				
Matches 166; Conservative 0; Mismatches 79; Indels 0; Gaps 0;				
QY	1	CTGCCCGCCAGACACAGCAACAACACACACCACCTTCCTCGCCCTGGTTCGCTCCCAC	60	
DB	26365	CTGCCCGCATGACACCGGACCGGATCGGCACCTTCGCGGACCTCGTCCGTACGCAC	26424	
QY	61	ATCGCCACCGTCTGGGCGCACACACCCCCGACACCATCCCCCGACCGCGTTCGCGC	120	
DB	26425	GTGGCGACCGTCTGGGACACCGGACCGGACCGGATCGGACCTGGAGCGGCGCTTCGCG	26484	
QY	121	GACTTCGGCTTCGACTCCCTCACCCTTGAACTAGCAACCGCTCTCCCGACACACC	180	
DB	26485	GACACCGGTTTCGACTCGCTCAGCGCGCTGAACTTCGCAACCGCTCTCAAACGCGGACC	26544	
QY	181	GGACTTCGCGCTCCCGACACACCTTCGCTTGACACACCCCAACCCACACCTTCACCCAC	240	
DB	26545	GGGCTGCGGCTGCGCGCACCGCTGCTTTCGACACACCCACCTTCGCGGAGCTCGCGGG	26604	
QY	241	CACCT 245		
DB	26605	CACCT 26609		
RESULT 7				
US-09-861-289-5				
; Sequence 5, Application US/09861289				
; Patent No. US20020110897A1				
; GENERAL INFORMATION:				
; APPLICANT: Sherman, D.H.				
; APPLICANT: Liu, H.				
; APPLICANT: Xue, Y.				
; APPLICANT: Zhao, L.				
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin				
; FILE REFERENCE: 600.438U51				
; CURRENT APPLICATION NUMBER: US/09/861.289				
; CURRENT FILING DATE: 2001-05-18				
; PRIOR APPLICATION NUMBER: 09/105.537				
; PRIOR FILING DATE: 1998-06-26				
; NUMBER OF SEQ ID NOS: 43				
; SOFTWARE: FastSeq for Windows Version 3.0				
; SEQ ID NO 5				
; LENGTH: 36778				
; TYPE: DNA				
; ORGANISM: Streptomyces venezuelae				
US-09-861-289-5				
Query Match 48.2%; Score 118.6; DB 10; Length 36778;				
Best Local Similarity 67.8%; Pred. No. 1.4e-20;				
Matches 166; Conservative 0; Mismatches 79; Indels 0; Gaps 0;				
QY	1	CTGCCCGCCAGACACAGCAACAACACACACCACCTTCCTCGCCCTGGTTCGCTCCCAC	60	
DB	26365	CTGCCCGCATGACACCGGACCGGATCGGCACCTTCGCGGACCTCGTCCGTACGCAC	26424	
QY	61	ATCGCCACCGTCTGGGCGCACACACCCCCGACACCATCCCCCGACCGCGTTCGCGC	120	
DB	26425	GTGGCGACCGTCTGGGACACCGGACCGGACCGGATCGGACCTGGAGCGGCGCTTCGCG	26484	
QY	121	GACTTCGGCTTCGACTCCCTCACCCTTGAACTAGCAACCGCTCTCCCGACACACC	180	
DB	26485	GACACCGGTTTCGACTCGCTCAGCGCGCTGAACTTCGCAACCGCTCTCAAACGCGGACC	26544	
QY	181	GGACTTCGCGCTCCCGACACACCTTCGCTTGACACACCCCAACCCACACCTTCACCCAC	240	
DB	26545	GGGCTGCGGCTGCGCGCACCGCTGCTTTCGACACACCCACCTTCGCGGAGCTCGCGGG	26604	
QY	241	CACCT 245		
DB	26605	CACCT 26609		
RESULT 6				
US-09-836-821-5				
; Sequence 5, Application US/09836821				
; Publication No. US20030087405A1				
; GENERAL INFORMATION:				
; APPLICANT: Sherman, D.H.				
; APPLICANT: Liu, H.				
; APPLICANT: Xue, Y.				
; APPLICANT: Zhao, L.				
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin				
; FILE REFERENCE: 600.438U51				
; CURRENT APPLICATION NUMBER: US/09/836.821				
; CURRENT FILING DATE: 2001-04-17				
; PRIOR APPLICATION NUMBER: 09/105.537				
; PRIOR FILING DATE: 1998-06-26				
; NUMBER OF SEQ ID NOS: 43				
; SOFTWARE: FastSeq for Windows Version 3.0				
; SEQ ID NO 5				
; LENGTH: 36778				
; TYPE: DNA				

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 29.0718 Seconds
(without alignments)
12253.192 Million call updates/sec

Title: US-09-914-286-1_COPY_11413_11658

Perfect score: 246

Sequence: 1 ctggcgccgacagacacga.....ccacctccaccacacaccc 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.6	48.2	11220	9	US-09-860-846-32
2	118.6	48.2	11220	9	US-09-988-384B-32
3	118.6	48.2	11220	9	US-09-836-821-32
4	118.6	48.2	11220	10	US-09-861-289-32
5	118.6	48.2	36778	9	US-09-860-846-5
6	118.6	48.2	36778	9	US-09-836-821-5
7	118.6	48.2	36778	10	US-09-861-289-5
8	118.6	48.2	37948	9	US-09-988-384B-5
9	118.6	48.2	38506	9	US-09-793-708-19
10	118.2	48.0	50937	9	US-09-808-880-1
11	111.2	45.2	4689	9	US-09-860-846-34
12	111.2	45.2	4689	9	US-09-988-384B-34
13	111.2	45.2	4689	9	US-09-836-821-34
14	111.2	45.2	4689	10	US-09-861-289-34
15	110	44.7	13842	9	US-09-860-846-30
16	110	44.7	13842	9	US-09-988-384B-30
17	110	44.7	13842	9	US-09-836-821-30
18	110	44.7	13842	10	US-09-861-289-30
19	103.6	42.1	15872	9	US-09-860-846-1

20 103.6 42.1 15872 9 US-09-988-384B-1 Sequence 1, Appli

21 103.6 42.1 15872 9 US-09-836-821-1 Sequence 1, Appli

22 103.6 42.1 15872 10 US-09-861-289-1 Sequence 1, Appli

23 95.6 38.9 4041 9 US-09-860-846-36 Sequence 36, Appl

24 95.6 38.9 4041 9 US-09-988-384B-36 Sequence 36, Appl

25 95.6 38.9 4041 9 US-09-836-821-36 Sequence 36, Appl

26 95.6 38.9 4041 10 US-09-861-289-36 Sequence 36, Appl

27 86 35.0 4851 9 US-09-712-363-116 Sequence 116, App

28 75.4 30.7 4209 9 US-09-712-363-20 Sequence 20, Appl

29 73.2 29.8 520 9 US-10-184-644-332 Sequence 332, App

30 73.2 29.8 520 9 US-10-184-644-332 Sequence 332, App

31 73 29.7 12733 9 US-10-032-393-47 Sequence 47, Appl

32 73 29.7 12733 9 US-10-032-393-8 Sequence 8, Appli

33 69.4 28.2 68750 9 US-10-014-717-1 Sequence 1, Appli

34 67 27.2 925 9 US-09-735-056-1 Sequence 1, Appli

35 67 27.2 152331 9 US-10-095-407-16 Sequence 16, Appl

36 61.8 25.1 1064 10 US-09-804-682-29 Sequence 29, Appl

37 60.8 24.7 504 10 US-09-878-574-4330 Sequence 4330, Ap

38 59.4 24.1 1065 10 US-09-804-682-33 Sequence 33, Appl

39 58.8 23.9 266 10 US-09-960-352-3882 Sequence 3882, Ap

40 57.6 23.4 2163 9 US-10-198-846-12766 Sequence 12766, A

41 57 23.2 437 10 US-09-960-352-12835 Sequence 12835, A

42 56.6 23.0 931 9 US-10-198-846-8585 Sequence 8585, Ap

43 56.2 22.8 987 10 US-09-804-682-20 Sequence 20, Appl

44 55.6 22.6 1194 10 US-09-804-682-7 Sequence 7, Appli

45 55.6 22.6 3453 9 US-10-101-464A-861 Sequence 861, App

ALIGNMENTS

RESULT 1

US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.438US1

; CURRENT APPLICATION NUMBER: US/09/860,846

; CURRENT FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: 09/105,537

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 32

; LENGTH: 11220

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

; US-09-860-846-32

Query Match 48.2%; Score 118.6; DB 9; Length 11220;
Best Local Similarity 67.8%; Pred. No. 1.7e-20;
Matches 166; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1 CTGCGCGCCAGACACACGACACACACCTCTCGCCCTCGCTCGCTCCAC 60
10678 CTCGCGCGATGACACACCGAGACGCGGTGCGGACCTCGTCCGTACGCAC 10737

QY 61 ATCCGCCCGCTCGGGCCACACCCACCATCCCCCGCGCGGCTTCGCG 120
10738 GTGCGGACCGCTCGGGACACCGCACCCCGCGGCTGGAGCGGCTTCGCG 10797

QY 121 GACCTCGGCTTCGACTCTCCTCACCGCCCTCGAACTACGCAACCGGCTCTCCGCGACACCC 180
10798 GACACCGGTTTCGACTCGCTCACCGCCCTCGAACTACGCAACCGGCTCTCAACCGCGCGACC 10857

QY 181 GGACTCGGCTTCGCGACACACCTCGCTTCGACACACCCACCCACCTTACCCAC 240
10858 GGGTTCGCGCTGCGCGCCACGCTGTCTTCGACACCCCGCGGGGAGTCTCGCCGGG 10917

THIS PAGE BLANK (USPTO)

APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 43.7%; Score 107.4; DB 3; Length 80161;
Best Similarity 64.9%; Pred. No. 4.3e-13;
Matches 159; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 1 CTGCGCGGCGGACACACGAAACAAGACACACACCTCTCGCGCTGTCCTCCAC 60
DB 39790 CTGCGGAGCTGCGGAGCGCAAGCCCAACGACGCGCAACTGATGATGATGCGCCAG 39849
QY 61 ATGCCACCGTCTCTGGGCGCACACACCCCGACACATCCCTCCCGACGCGGTTCCG 120
DB 39850 GTGGCAACCGTCTCGGCGCAGCGCAGTCGAGAGAACTCCAGCCCGCGGCTTCGCG 39909
QY 121 GACCTGGGCTTTCGACTCCCTCACCGCGGTGAACTACGAAACGGGCTCTCCGACACAC 180
DB 39910 GCGCTGGGTTTCGACTCTCTCAAGCGGTGATCTGGCAATGTTGACACCGCCACC 39969
QY 181 GGAATCGGCTCCCGACACACCTTCGCTTGAACACACCCCAACCTTCACCCAC 240
DB 39970 GGGTTGGGCTCGCGACACCAACCGCTTTCGACTACCCGGAATCGGCGGCTTGGCGCT 40029
QY 241 CACCT 245
DB 40030 CACCT 40034

Search completed: June 18, 2003, 01:08:30
Job time : 26.6565 secs

```
OTHER INFORMATION: /function= "approximate span of
FEATURE: module 3"
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPSynthase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
```

```
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19482..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3
Query Match 44.2% Score 108.8; DB 3; Length 20235;
Best Local Similarity 70.2%; Pred. No. 2,3e-13;
Matches 146; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 36 CCTCTGCGCCCTGTGCTCGCTCCACATGCGCACCGTCTGTGGGCGCACACCCCGGACAC
DB 14893 CCTGCCACCTGATCTCGGCGGAGGTGCGCGGCTGCTGGGCCACGGGACGAGCGGCG
QY 96 CATCCCCCGGACCGCGCGGTTCCGGACCTTGAGCTTCACTTCCCTACCGCGCTGAACT
DB 14952 GATTCGACCGCGACCGCGGCTTCCGGACCTTGAGCTTCACTTCCACGACCGCGTGACT
QY 156 AGCGAAGCGGCTCTCCGCGACCGAGGACTCGGCTCCGACACCACTTGACTTGACCA
DB 15012 GCGGAACCGGCTTCCGCGGCGGAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 216 CCCGAACCGCACCACTTACCCACAC 243
DB 15072 CCCGACATCAACCGGCTGCGGACAC 15099
RESULT 15
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
```

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACpynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACpynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19145..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: chioesterase domain of module 6"
US-07-642-734C-3

```

```

Query Match      44.2%; Score 108.8; DB 1; Length 20235;
Best Local Similarity 70.2%; Pred. No. 2.3e-13;
Matches 146; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

Qy 36 CCTCTGCGCCCTGTGCTCGCTCCGACATGCGCACCGTCTTGCGGCACACCAACCCCGGACAC
Db 14892 CCTGCGCCACCTGATCGCGCGGAGGTGCGCGGCTGTGCGGCACGCGGAGCGCGGCG 14951
Qy 96 CATCCCCCCCCGACCGGCGGTTCGCGGACCTCGGCTTGCACTCTCTACCGCGGTGCAACT 155
Db 14952 GATCGACCGCGACCGGCGCTTCGCGGACCTCGGCTTGCACTCTACGACCGCGTCACTT 15011
Qy 156 AGCAACCGGCTCTCCGCGACCGGACCGGACTCGGCTCCGACACACCGTTCGACCA 215
Db 15012 GCGGACCGGCTGCGCGCGGTGACCGGCGGTGCGGGAAGCTCGGACGCGGTGCTTGACCA 15071
Qy 216 CCCCAACCCGACCGGCTTCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 243
Db 15072 CCGGACGATCACCGGCGCTGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 15099

```

```

RESULT 14
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, J
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstock
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439, 009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722
; OTHER INFORMATION: /codon_start= 19
; OTHER INFORMATION: /function= "gene =eryA"
; OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19..4470

```


TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 52.1%; Score 128.2; DB 2; Length 43280;
Best Local Similarity 70.2%; Pred. No. 4.4e-17;
Matches 172; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 CTGGCCGCGCAGACACAGACACACACACACACACCTCTGCGCTGCTCCGAC 60
DB 30648 CTGACCGGCGCTCAGCAGACAGAGAGGCGACGCTCTGCTGCGAGATCCGCGCCAC 30707
QY 61 ATCGCACCGCTCTGGGCGCACACACCCCGACACCATCCCGCGCGGCTTCCG 120
DB 30708 GCGGCGCGCGTGTGGGCGCACGCGCTCGACGACTGATCCCGAGAGCGGCGCTTCAAG 30767
QY 121 GACCTCGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCCGACACAC 180
DB 30768 GACCTCGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCCGACACAC 30827
QY 181 GACCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTCAACCCAC 240
DB 30828 GGCCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTGAACCCG 30887
QY 241 CACCT 245
DB 30888 CACCT 30892

RESULT 4
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 50.2%; Score 123.6; DB 4; Length 4403765;
Best Local Similarity 69.4%; Pred. No. 3.4e-16;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CTGGCCGCGCAGACACAGACACACACACACACCTCTGCGCTGCTCCGAC 60
DB 2297332 CTGGAAGAGCTGCCGGAAGACAGACAGCGCTCTCTGACTGTGCGCTCGAC 2297273
QY 61 ATCGCACCGCTCTGGGCGCACACACCCCGACACCATCCCGCGCGCGCTTCCG 120
DB 2297272 ATCGCACCGCTCTGGGCGCACACCCCGACACCATCCCGCGCGCTTCCG 2297213
QY 121 GACCTCGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCCGACACAC 180
DB 2297212 GAGTTGGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCAATCCG 2297153
QY 181 GACCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTCAACCCAC 240
DB 2297152 GGCCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTGAACCCG 2297093
QY 241 CA 242
DB 2297092 TA 2297091

RESULT 5
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 50.2%; Score 123.6; DB 4; Length 4411529;
Best Local Similarity 69.4%; Pred. No. 3.4e-16;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CTGGCCGCGCAGACACAGACACACACACACCTCTGCGCTGCTCCGAC 60
DB 2295032 CTGGAAGAGCTGCCGGAAGACAGACAGCGCTCTCTGACTGTGCGCTCGAC 2294973
QY 61 ATCGCACCGCTCTGGGCGCACACACCCCGACACCATCCCGCGCGCGCTTCCG 120
DB 2294972 ATCGCACCGCTCTGGGCGCACACCCCGACACCATCCCGCGCGCTTCCG 2294913
QY 121 GACCTCGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCCGACACAC 180
DB 2294912 GAGTTGGGCTTGACTCTCTCAACCGCGCTGAACTAGCAACCGGCTCTCAATCCG 2294853
QY 181 GACCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTCAACCCAC 240
DB 2294852 GGCCTCGGCTTGACTCTCTCAACCGCGCTTGAACACACCCGACACACCTTGAACCCG 2294793
QY 241 CA 242

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 10.6565 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658

Perfect score: 246
Sequence: 1 CTGCGCCGCGCAGACACACGA.....CCACCTCACACACACCTC 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfillseq.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137.2	55.8	44377	2	US-08-804-227C-7
2	137.2	55.8	44377	2	US-08-804-198-1
3	128.2	52.1	43280	2	US-08-804-227C-1
4	123.6	50.2	4403765	4	US-09-103-840A-2
5	123.6	50.2	4411529	4	US-09-103-840A-1
6	121.8	49.5	1681	4	US-09-434-288-7
7	118.6	48.2	11220	4	US-09-105-537-32
8	118.6	48.2	36778	4	US-09-105-537-5
9	118.6	48.2	38506	3	US-09-320-878-19
10	118.2	48.0	50937	4	US-09-428-517-1
11	111.2	45.2	4689	4	US-09-105-537-34
12	110	44.7	13842	4	US-09-105-537-30
13	108.8	44.2	20235	1	US-07-642-734C-3
14	108.8	44.2	20235	1	US-08-439-009A-3
15	107.4	43.7	80161	3	US-09-336-987A-1
16	107.4	43.7	80161	3	US-09-370-700-1
17	105.2	42.8	1434	4	US-09-434-288-3
18	104.4	42.4	13987	2	US-08-804-227C-13
19	103.6	42.1	15872	4	US-09-105-537-1
20	102	41.5	49377	1	US-08-764-233A-1
21	100.4	40.8	11219	3	US-08-439-009A-1
22	100.4	40.8	11219	3	US-07-642-734C-1
23	99.2	40.3	28958	1	US-08-456-837-6
24	99.2	40.3	28958	1	US-08-456-837-6
25	99.2	40.3	28958	1	US-08-457-342-6
26	99.2	40.3	28958	1	US-08-457-646A-6
27	99.2	40.3	28958	1	US-08-458-076A-6

28	99.2	40.3	28958	1	US-08-764-233A-4	Sequence 4, Appli
29	99.2	40.3	28958	1	US-08-457-335A-6	Sequence 6, Appli
30	99.2	40.3	28958	1	US-08-729-214-6	Sequence 6, Appli
31	99.2	40.3	28958	3	US-09-028-934-6	Sequence 6, Appli
32	95.6	38.9	4041	4	US-09-105-537-36	Sequence 36, Appli
33	87.6	35.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
34	87.6	35.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
35	86.2	35.0	561	3	US-09-154-083-13	Sequence 13, Appli
36	78.6	32.0	748	3	US-09-154-083-1	Sequence 1, Appli
37	78.6	32.0	753	3	US-09-154-083-28	Sequence 28, Appli
38	77.6	31.5	318	4	US-09-165-264-12	Sequence 12, Appli
39	74.4	30.2	320	4	US-09-165-264-14	Sequence 14, Appli
40	73.8	30.0	319	4	US-09-165-264-8	Sequence 8, Appli
41	73.8	29.9	320	4	US-09-165-264-7	Sequence 7, Appli
42	73.6	29.9	320	4	US-09-165-264-13	Sequence 13, Appli
43	71.6	29.8	320	4	US-09-165-264-11	Sequence 11, Appli
44	71.6	29.1	33529	4	US-09-144-085-3	Sequence 3, Appli
45	70.6	28.7	367	4	US-09-060-756-627	Sequence 627, App

ALIGNMENTS

RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:

PA (STRO/) STROM A R.
 XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX WPI; 2001-557614/62.
 DR P-PSDB; AAE10153.
 XX New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX Example 2; Page 186-187; 266pp; English.
 XX The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a ERD48 insert DNA in pKR48 which is used in an
 CC exemplification of the invention.
 XX
 SQ Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;
 Query Match 50.2%; Score 123.4; DB 22; Length 2700;
 Best Local Similarity 73.7%; Pred. No. 7.9e-13;
 Matches 157; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 QY 33 CACCCTCTCGCCCTGCTCGCTCCACATCGCCACCGTCTGGGGCCACACACCCCGA 92
 DB CACCTCTCTCGGCTGCTCGGACACAGCGCGGATGGTCTCGGCCACACCTCGGGCGA 2065
 QY 93 CACCATCCCCCGACCGCGGTTCGCGACCTCGGCTTCGACTCCCTCACCAGCGTCCA 152
 DB CGCGCTCGACCGCTCCCGGCTTCGCGACCTCGGCTTCGACTCGCTCACCAGCGTCCA 2125
 QY 153 ACTACGCAACCGCTCTCCGACACACCGGACTCCGCTCCACACACCTCGGCTTCCA 212
 DB 2126 ACTCCGCAACCGGATCGGCGCGCCACCGGCTTCGCGCTACCGCCACCGCGCTTCCA 2185
 QY 213 CACACCCCAACCCACACACCTCTACCCACACACCT 245
 DB 2186 CTACCCCGCGCGATGCTCTCGCGCACACCT 2218

RESULT 15
 AAS18438
 ID AAS18438 standard; DNA; 1681 BP.
 XX
 AC AAS18438;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Contig 115 DNA encoding S. narbonensis polyketide synthase.
 XX
 KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
 KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethylmycin;
 KW agriculture; ds.
 XX
 OS Streptomyces narbonensis.
 XX
 PN US6303767-B1.
 XX
 PD 16-OCT-2001.
 XX
 PF 05-NOV-1999; 99US-0434288.
 XX
 PR 05-NOV-1998; 98US-107093P.
 XX 27-MAY-1999; 99US-0320878.
 XX
 PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX
 PI Betlach MC, McDaniel R;
 XX WPI; 2002-065495/09.

XX Nucleic acids encoding narbonolide polyketide synthases from
 PT Streptomyces narbonensis, useful for the recombinant production of
 PT polyketides, e.g. narbomycin -
 XX Claim 1; Column 18; 24pp; English.
 XX
 CC The present invention relates to recombinant DNA vectors (cosmids)
 CC that encode for the narbonolide polyketide synthase (PKS) enzyme and
 CC various narbomycin modification enzymes from Streptomyces narbonensis.
 CC The recombinant DNA vectors can be used to produce recombinant ketide
 CC synthases and a variety of different polyketides (e.g. erythromycin,
 CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
 CC neomethylmycin) for use in agriculture, medicine and health. The
 CC recombinant vectors may be used to produce polyketides in relatively
 CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
 CC encode for S. narbonensis PKS enzymes.
 XX
 SQ Sequence 1681 BP; 231 A; 672 C; 580 G; 198 T; 0 other;

Query Match 49.5%; Score 121.8; DB 24; Length 1681;
 Best Local Similarity 68.6%; Pred. No. 1.6e-12;
 Matches 168; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
 QY 1 CTGGCCCGCCAGACACACGAAACAGACACACCTCTCTCGCCCTCGGTCCGTCAC 60
 DB 541 CTGCGCGCGATGACCCCGACGACAGGACCGGACCTGCGGAGCTGTCGTACGCAC 600
 QY 61 ATCGCCACCGTCTGGGGCCACACACCCCGACACCATCCCCCGACCGCGGCTTCGCG 120
 DB 601 GTGGCGACCGTCTGGGACACGCGGCCCGGCGGCTCGACCTGGAGCGGCGCTTCGCG 660
 QY 121 GACCTCGGCTTCGACTCCCTCACCAGCGCTCGAACTACGCAACCGGCTCTCCCGCACCCACC 180
 DB 661 GACACCGGTTTCGACTCCCTCACCAGCGCTCGAGCTCCGCAACCGGCTCAACCGCGCCACC 720
 QY 181 GGACTTCGCGCTCCCGCCACACCCCTCGCCCTTCGACACACCCCAACCCCAACCCCTCAACCCAC 240
 DB 721 GGCGTTCGCGCTCCCGCCACGCTGCTTCGACACACCCCACTCCCGGGGAGCTCGCGCGG 780
 QY 241 CACCT 245
 DB 781 CACCT 785

Search completed: June 17, 2003, 13:49:16
 Job time : 61.0269 secs

Db 26912 GGCCTGCACCTGCGGCCACCATGGTCTTCGACACCCACCCCTGTCGCGCCCTCGCGGAG 26971
QY 241 CACCT 245
Db 26972 CACCT 26976

RESULT 11
AAT80413
XX ID AAT80413 standard; DNA; 43280 BP.
XX AC AAT80413;
XX DT 27-FEB-1998 (first entry)
XX DE Ty lactone synthase gene cluster.
XX KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
XX OS Streptomyces fradiae.
XX PH Key
XX FT Location/Qualifiers
XX FT 816..14243
XX FT /*tag= a
XX FT /transl_except= (pos: 816..818, aa: Met)
XX FT /note= "ORF1 encodes protein shown in AAW22601"
XX FT CDS
XX FT 14351..19945
XX FT /*tag= b
XX FT /transl_except= (pos: 14351..14353, aa: Met)
XX FT /note= "ORF2 encodes protein shown in AAW22602"
XX FT CDS
XX FT 20010..31199
XX FT /*tag= c
XX FT /transl_except= (pos: 20010..20012, aa: Met)
XX FT /note= "ORF3 encodes protein shown in AAW22603"
XX FT CDS
XX FT 31232..36067
XX FT /*tag= d
XX FT /note= "ORF4 encodes protein shown in AAW22604"
XX FT CDS
XX FT 36249..41774
XX FT /*tag= e
XX FT /note= "ORF5 encodes protein shown in AAW22605"
XX PN EP791655-A2.
XX XX
XX PD 27-AUG-1997.
XX XX
XX PP 19-FEB-1997; 97EP-0301056.
XX XX
XX PR 22-FEB-1996; 96US-0012078.
XX XX
XX PA (ELIL) LILLY & CO ELI.
XX XX
XX PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
XX XX
XX DR WPI; 1997-418046/39.
XX DR P-PSDB; AAW22601-W22605.
XX XX
XX PT DNA encoding Streptomyces fradiae ty lactone synthase domain - for
XX PT production of tylosin-related polyketide compounds
XX XX
XX PS Claim 2; Pages 8-66; 220pp; English.
XX XX
XX CC This sequence represents the ty lactone synthase gene cluster of the
XX CC invention. This sequence is also referred to as the tylG gene, and was
XX CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
XX CC proteins which direct the synthesis of the polyketide ty lactone, isolated
XX CC from Streptomyces fradiae. Ty lactone is the basic building block of the
XX CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
XX CC type of carboxylic acids incorporated, the number of carboxylic acids
XX CC incorporated and/or the post-condensation reactions performed, thereby
XX CC resulting in novel tylosin-related polyketides.
XX XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

Query Match 52.1%; Score 128.2; DB 18; Length 43280;
Best Local Similarity 70.2%; Pred. No. 8.6e-14;
Matches 172; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 1 CTGGCGCGCCAGACACAGCAACAGCAGCACCCTCTCTCGCCCTGGCTCCGCTCCAC 60
Db 30648 CTGACCGGGCTTCACGGCAGAGAGGGCAGCCCTCTCTGGCGGAGATCCGGGCCAC 30707
QY 61 ATCGCCACCGCTCTGGGCGCACACACCCCGACACCATCCCGCGCGCGCTTCGCG 120
Db 30708 CGCGGGGGTGTCTGGGCCACCGGTCTCGACACCTCGATCCCGAGGACCGGGCTTCAAG 30767
QY 121 GACCTCGGCTTTCGACTCTCTCACCGCGCTCGAATCTAGCAACCGGCTCTCCCGCACACC 180
Db 30768 GACCTCGGCTTTCGACTCTCTCACCGCGGTGAGATGGCAACCGGCTGAGCGCGGCCACC 30827
QY 181 GGATCGCGCTTCCCGCACACCGCTCTCGACACCCCAACCCACCCCTCACCCAC 240
Db 30828 GGCCTCGGCTTCCCGCGCACCCCTCTCTTGACACCCCGCGCGCGGAGCTGGCCGGC 30887
QY 241 CACCT 245
Db 30888 CACCT 30892

RESULT 12
AAI99683/c
ID AAI99683 standard; DNA; 4403765 BP.
XX AC AAI99683;
XX DT 15-JAN-2002 (first entry)
XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX OS Mycobacterium tuberculosis.
XX PN US6294328-B1.
XX XX
XX PD 25-SEP-2001.
XX XX
XX PF 24-JUN-1998; 98US-0103840.
XX XX
XX PR 24-JUN-1998; 98US-0103840.
XX XX
XX PA (GENO-) INST GENOMIC RES.
XX XX
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX XX
XX DR WPI; 2001-647261/74.
XX XX
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX PT determining the nucleotide sequence of the strain at positions in the
XX PT genome corresponding to positions where M. tuberculosis strains CDC
XX PT 1551 and H37Rv differ -
XX XX
XX PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX XX
XX CC The invention relates to evaluating strain variation within and between
XX CC different populations of the tuberculosis bacterial pathogen.
XX CC Mycobacterium tuberculosis or related Mycobacterium by determining the
XX CC nucleotide sequence of the first strain at positions that differ in the
XX CC sequence of the genome that correspond to positions that differ in the
XX CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
XX CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
XX CC M. tuberculosis and has valuable application in the fields of
XX CC tuberculosis genetics, epidemiology, patient treatment and epidemic
XX CC monitoring.
XX CC Note: The sequence data for this patent did not form part of the printed

FT		/product= "NysR4 (short) protein"	
FT		/note= "CDS does not include start codon"	
FT	CDS	61736..62497	
FT		/*tag= k	
FT		/product= "NysR5 protein"	
FT		/note= "CDS does not include start codon"	
FT	CDS	complement (62551..63615)	
FT		/*tag= l	
FT		/product= "ORF2 protein"	
FT		/note= "CDS does not include start codon"	
FT	CDS	63765..64961	
FT		/*tag= m	
FT		/product= "ORF1 protein"	
XX	XX		
XX	PD	WO200159126-A2.	
XX	PD	16-AUG-2001.	
XX	XX		
PF	08-FEB-2001;	2001WO-GB00509.	
XX	XX		
PR	08-FEB-2000;	2000GB-0002840.	
PR	10-APR-2000;	2000GB-0008786.	
PR	14-APR-2000;	2000GB-0009387.	
XX	XX		
PA	(UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.		
PA	(SNTF) SINTEF STIFTELSEN IND TEK FORSK.		
PA	(ALPH-) ALPHARMA AS.		
PA	(SINV-) SINVENT AS.		
PA	(DZIE/) DZIEGLEWSKA H.		
PA	(ZOTC/) ZOTCHEV S B.		
PA	(SEKU/) SEKUROVA O N.		
PA	(FJAE/) FJAEVRIK E.		
PA	(BRAU/) BRAUTASET T.		
PA	(STRO/) STROM A R.		
XX	XX		
FI	Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;		
FI	Valla S, Ellingsen TE, Sletta H, Gulliksen O;		
XX	XX		
DR	WPI; 2001-557614/62.		
DR	P-PSDB: AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130.		
DR	AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.		
XX	XX		
PT	New nystatin polyketide synthase polynucleotides and polypeptides,		
PT	useful as antibiotics and antifungals -		
XX	XX		
PS	Claim 2; Page 116-151; 266pp; English.		
XX	XX		
CC	The present invention relates to the cloning and sequencing of the gene		
CC	cluster encoding a modular type I polyketide synthase (PKS) enzyme		
CC	involved in the biosynthesis of the macrolide antibiotic nystatin.		
CC	The nystatin PKS is useful as antifungal antibiotics. The present		
CC	sequence is a Streptomyces noursei nyl DNA of nystatin PKS gene cluster.		
XX	XX		
SQ	Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;		
	Query Match	52.8%; Score 129.8; DB 22; Length 65140;	
	Best Local Similarity	70.6%; Pred. No. 4.4e-14;	
	Matches 173; Conservative	0; Mismatches 72; Indels 0; Gaps 0;	
QY	1 CTGCGCGGCAGACACACGAACAACAGCACACCCCTCTCGCCCTGGTCCGCTCCAC 60		
Db	26732 CTGCGCGGACTCGACCACGCCCAACGCCACGAGGACACTGCTCGATCATGTGCGCAGCAGC 26791		
QY	61 ATCGCCACCGTCTGGCGCACACACCACCCCGACACATCCCCCGACGCGCGGTTCGC 120		
Db	26792 GCGCGCTCGTCTCGGCCACGCCGACGCGCAACGCGCATCGACGCCGAACGCGCTTCGC 26851		
QY	121 GACCTCGGCTTGACACTCCCTCAACCGCGTCGCAACTACGCAACCGGCTCTCCGCAACCCAC 180		
Db	26852 GACCTCGGCTTGACACTCGTCAACCGGCTTGAACTCCGCAACCGTCTGGCAGCCGACCC 26911		
QY	181 GGACTCGGCTTCCCAACACACCTCGCTTCGACACCCCAACCCCAACCCCTCACCAC 240		

QY 1 CTGGCGGCGGAGACACAGCAACAGACACACCTCTCTCGCCCTCGTGTTCGCTCCAC 60
|||||
Db 12230 CTGGCGGCGGAGACACAGCAACAGCAACACACCTCTCTCGCCCTCGTGTTCGCTCCAC 12289
|||||
QY 61 ATCGCCACCGTCTGGGCGACACACCCCGGACACCATCCCGCCGACCGCGGTTCGCG 120
|||||
Db 12290 ATCGCCACCGTCTGGGCGACACACCCCGGACACCATCCCGCCGACCGCGGTTCGCG 12349
|||||
QY 121 GACCTCGGCTTCTGACTCTCCCTCACCGCGGTGCA 152
|||||
Db 12350 GACCTCGGCTTCTGACTCTCCCTCACCGCGGTGCA 12381
|||||

RESULT 7

AAT78508
ID AAT78508 standard; DNA; 44377 BP.

XX AC AAT78508;
XX DT 26-FEB-1998 (first entry)
XX DE Platenolide synthase gene cluster.

XX KW Platenolide synthase gene cluster; platenolide production; srmG gene;
XX KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX OS Streptomycetes ambofaciens.

XX FH Key Location/Qualifiers
XX FT CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAW23716"

FT CDS 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAW23717"

FT CDS 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAW23718"

FT CDS 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAW23719"

FT CDS 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAW23720"

XX EP791656-A2.
XX PD 27-AUG-1997.

XX PF 19-FEB-1997; 97EP-0301066.

XX PR 22-FEB-1996; 96US-0012050.

XX PA (ELIL) LILLY & CO ELI.

XX PI Burgett SG, Kuhstoss SA, Rao RN, Richardson MA;
PI Rosteck PR;

XX WPI; 1997-418047/39.

XX DR P-PSDB; AAW23716-W23720.

XX PT DNA encoding Streptomycetes ambofaciens platenolide synthase domain -
XX PT for production of spiramycin-related polyketide antibiotics
XX PS Claim 9; Pages 8-33; 81pp; English.

XX CC This sequence represents the platenolide synthase gene cluster of the
XX CC invention. This sequence is referred to as the srmG gene, and was
XX CC isolated from Streptomycetes ambofaciens. This sequence encodes the
XX CC multi-functional proteins which direct the synthesis of the polyketide
XX CC platenolide. Platenolide is the basic building block of the macrolide

CC antibiotic spiramycin. The DNA can be used to produce compounds
CC exhibiting antibiotic activity based on the platenolide structure,
CC including specifically the macrolide antibiotic spiramycin and spiramycin
CC analogues and derivatives. Modifications of the platenolide synthase DNA
CC sequence can be made so as to change the number and type of carboxylic
CC acids incorporated into the growing polyketide chain and to change the
CC kind of post-condensation processing that is conducted.

XX SQ Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

Query Match 55.8%; Score 137.2; DB 18; Length 44377;
Best Local Similarity 72.4%; Pred. No. 2.6e-15;
Matches 178; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1 CTGGCGGCGGAGACACAGCAACAGACACACCTCTCTCGCCCTCGTGTTCGCTCCAC 60
|||||

Db 24427 CTGACGCGCGAGTCTGGGCACAGCGGACGCGCACCTCTCTCGGCTGTGTGGCGCACAC 24486
|||||

QY 61 ATCGCCACCGTCTCTGGGCGACACACCCCGGACACCATCCCGCCGACCGCGGTTCGCG 120
|||||

Db 24487 GCGGCGGCGTCTCTGGGACAGTCTCTCGGCGAGCGGTGAGCAGCGCGCTTCGCG 24546
|||||

QY 121 GACCTCGGCTTCTGACTCTCCCTCACCGCGGTTCGAACTACGCAACCGGTCTCTCGCACAC 180
|||||

Db 24547 GACCTCGGCTTCTGACTCTCCCTCACCGCGGTTCGAACTACGCAACCGGTCTCTCGCACAC 24606
|||||

QY 181 GGACTCCGCTCTCCCGACACCTCTCGCTTCGACACACCCCAACCCACACCTTCACCCAC 240
|||||

Db 24607 GGGCTCAAACTGCGCCACCTCTCTGTCTTCGACCACTTCAGCCCGCGCTTCGCGCGG 24666
|||||

QY 241 CACCTC 246

Db 24667 CACCTC 24672

RESULT 8

AAT80414

ID AAT80414 standard; DNA; 44377 BP.

XX AC AAT80414;
XX DT 27-FEB-1998 (first entry)
XX DE Platenolide synthase gene cluster.

XX KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
XX KW platenolide synthase gene cluster; platenolide production; srmG gene;
XX KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.

XX OS Streptomycetes ambofaciens.

XX FH Key Location/Qualifiers
XX FT CDS 350..14002
FT /tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAW22606"

FT CDS 14046..20036
FT /tag= b
FT /note= "ORF2 encodes protein shown in AAW22607"

FT CDS 20110..31284
FT /tag= c
FT /transl_except= (pos:20111..20113, aa:Met)

FT CDS 31329..36071
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAW22609"

FT CDS 36155..41830
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAW22610"

XX EP791655-A2.
XX PD 27-AUG-1997.

Db 20269 GGACTCGGCTCCCAACCACTTGGCTTTCGACCAACCCCAACCAACCACTTCCCGC 20328
QY 241 CACCTC 246
Db 20329 CACCTC 20334
RESULT 5
AAH79278
ID AAH79278 standard; DNA; 31422 BP.
XX
AC AAH79278;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO.: 2.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; veterinary drug; pesticide; db.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..14646
FT /*tag= a
FT /product= "AAG65266"
FT CDS 14824..31422
FT /*tag= b
FT /product= "AAG65267"
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
WPI; 2001-582053/65.
DR P-PSDB; AAG65266, AAG65267.
XX
XX New modified avermectin aglycone synthase derived from Streptomyces
PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Disclosure; Page 103-149; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermitilis genome.
XX
SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
Query Match 99.3%; Score 244.4; DB 22; Length 31422;
Best Local Similarity 99.6%; Pred. No. 2.8e-33;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTGGCGCGCGAGACACAGACAGACACCACTTCTCGCCCTGCTCGCTCCAC 60
Db 20089 CTGGCGCGCGAGACACAGACAGACACCACTTCTCGCCCTGCTCGCTCCAC 20148
QY 61 ATGCGCACCGTCTCTGGGCGACACCAACCCCGACACCATCCCTCCCGCGGTTCCGC 120

Db 20149 ATGCGCACCGTCTCTGGGCGACACCAACCCCGACACCATCCCTCCCGCGGTTCCGC 20208
QY 121 GACCTCGGCTTCGACTCCCTCAGCGCGTTCGAACTACGCAACCGGCTTCGGGACCAACC 180
Db 20209 GACCTCGGCTTCGACTCCCTCAGCGCGTTCGAACTACGCAACCGGCTTCGGGACCAACC 20268
QY 181 GGACTCGGCTTCGCGCCACCACTTGGCTTTCGACCAACCCCAACCACTTCCCGC 240
Db 20269 GGACTCGGCTTCGCGCCACCACTTGGCTTTCGACCAACCCCAACCACTTCCCGC 20328
QY 241 CACCTC 246
Db 20329 CACCTC 20334
RESULT 6
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
WPI; 2000-182117/16.
XX
XX Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 61.8%; Score 152; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1e-17;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 48.936 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658

Perfect score: 246

Sequence: 1 ctggcgccgacagacagca.....ccaccctcaccaccacccctc 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	100.0	11316	22	AAH79279
2	246	100.0	30690	21	AAA92301
3	246	100.0	30690	22	AAH79277
4	244.4	99.3	31422	21	AAA92302
5	244.4	99.3	31422	22	AAH79278
6	152	61.8	12381	21	AAZ59381
7	137.2	55.8	44377	18	AAT78508
8	137.2	55.8	44377	18	AAT80414
9	133	54.1	125401	22	AAD17186
					Streptomyces nours
					Streptomyces averm
					S. avermitilis ave
					Streptomyces averm
					S. avermitilis ave
					Streptomyces averm
					Streptomyces averm
					Platenolide syncha
					Platenolide syncha
					Streptomyces nours

10	129.8	52.8	65140	22	AAD17184
11	128.2	52.1	43280	18	AAT80413
c 12	123.6	50.2	4403765	22	AAI99683
c 13	123.6	50.2	4411529	22	AAI99682
14	123.4	50.2	2700	22	AAD17193
15	121.8	49.5	1681	24	AAS18438
16	118.6	48.2	11220	21	AZ87298
17	118.6	48.2	36778	21	AZ87318
18	118.6	48.2	37948	21	AZ87285
19	118.6	48.2	38506	21	AAZ75633
20	118.6	48.2	38506	21	AAZ56001
21	118.2	48.0	50937	21	AAO9469
22	118	48.0	27541	22	AAD17185
23	117	47.6	20394	22	AAF24892
24	112.2	45.6	53789	19	AAV21187
25	111.2	45.2	4689	21	AZ87299
26	110	44.7	13842	21	AZ87297
27	108.8	44.2	29879	14	AAQ46806
c 28	107.6	43.7	77536	21	AAAL4651
29	107.4	43.7	9513	22	AAF88337
30	107.4	43.7	47981	22	AAF30757
31	107.4	43.7	50000	22	AAF88313
32	107.4	43.7	50000	22	AAF88316
33	107.4	43.7	80161	20	AZ21501
c 34	105.2	42.8	1434	24	AAS18434
35	104.4	42.4	13987	18	AAT80415
36	103.6	42.1	15872	21	AZ87283
37	102	41.5	3978	21	AAC55785
38	102	41.5	12249	21	AAC55840
39	102	41.5	18331	21	AAC55857
40	102	41.5	49377	19	AAV05287
41	99.2	40.3	28598	17	AAT06769
42	99.2	40.3	28958	18	AAT89956
43	99.2	40.3	28958	21	AAZ75299
44	98.8	40.2	15872	18	AAT68715
45	97.8	39.8	5676	19	AAV21186

ALIGNMENTS

RESULT 1

AAH79279
ID AAH79279 standard; DNA; 11916 BP.

AC AAH79279;

XX
DT 04-DEC-2001 (first entry)

XX Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.

XX Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide; ds.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 1..11916

FT /*tag= a

FT /product= "AAG65268"

FT /partial

XX WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP01381.

XX 24-FEB-2000; 2000JP-0047405.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX (KITA) KITASATO INST.

XX


```

/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"
/protein_id="AAB66505.1"
/db_xref="GI.2317921"

```

Query Match 52.1%; Score 128.2; DB 1; Length 43280;
Best Local Similarity 70.2%; Pred. No. 7.6e-08;
Matches 172; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

61 ATGCGCACCGTCTGGGGCCACACCCCGGACACCATCCCCCGACGCGGCTTCGC 120
30708 GCGGGGGCGGTGCTGGGGCCACGGCTCGACGACTCGATCCCCGAGGACCGGGGCTTCAAG 30767

181 GGACTCGGCTCCGCCACCAACCCTCGCCTTCGACCACCCCACCCACCCCTCACCCAC 240
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30828 GGCCTCGGGTGC CGCGCCAC CCGCTGTCTTCGACCACCCCGGGCGAGCTGGCGGC 30887

RESULT 15
007061/C

MYCOTUB
AE007061.1 GI:13881778

1 (bases 1 to 15970)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolonyak,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished

2 (bases 1 to 15970)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolonyak,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
Gill,J., Mikula,A. and Bishai,W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1. .15970
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
70. .726
/gene="MT2106"
70. .726
/gene="MT2106"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK46385.1"
/db_xref="GI:13881779"
/translation="MRIALVAVSLIACGSEVGDVGQSQTIAPPAPAPSAAPT
PAAGATPTIVSMIEAGHPDPAAYHVAATRDGVTQLGDDVAFSSSGTVACWMDLH
TSGTACLVRLANPPRPETAYCEWKGWQVDFGTHLQVGSARADPGPVYGVNGPELA
NGDTLSIGDYRCRSYQAGLFCVYAHQSAVRFAAGIEPFGCLKPAPPDPDGVGVA
" complement (763. .3327)
/gene="MT2107"
complement (763. .3327)
/gene="MT2107"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK46386.1"
/db_xref="GI:13881780"
/translation="MRIATVGASVGLRGTLTARLLSQGHEVVGIARHRPDSWPSADF
TAADIRDTAVESAMTADVAACHCAWVRGNHDINIDCTANVLKAMAETGTGRIVFTS
SGHQPRVEQMLADCGLEWAVKCALI RGNVDNVQORLPALPULPAGYADRVVQVHS
DDAQLVRALLDTVIDSGPNNLAAPGLTFRRIAAALGRMPVIGSPVLRVRTVSFEE
LELLSHAPMLDTLLDRWGFQPAWNAEECLDFTLAVRGRIGLGKRTLSLWRLANI
QDLPAVSDPADGGA PRLAGPEGANGFDTPIDPRFTYLATNLSEALPSPGSPSSAS
TVVRLGAGGVGIAERLPSGVIQREIARTVAVFAHRLYGAITSAHFWAATVPFAKPR
ATIVNSGFGPFGMSALPIPGAQRPSSSRARWLRLTNRTGVGVNLVGLSAGSPR
TDDAVDVDRLEAFONLATHDDRRLLSLILLARDHVHGVWLAGSGFMLCAAFNV
LRLGCLGRDTPAAGPELVASRSEAVQALVAAARDPVVIRLLAEPPGERLKLAVEA
PFHSAVAELTILGHREPAEYEMAATSYADNPPELLVRMAKTLRAVPAQPTTPIV
PRKFAVLLAAQLRDREVRDRMRAVIRALLREYGRRLTEAGVFTDTPDDVFVLL
VDEIDALPADVSLVARRAEORRLAGIVPPTVFGSGMEPPSSAAALAAAGDTLGRGV
VGVGRVGRVIRVIRETDDQLQPGELVAIVTVDVGYTAAFCYAAAVVTELGPMASHA
VVARFPGPCVVDAQGAFTRFPGFALVEVDGATGTHVVEVLASEDGPALFGGSLSR"
complement (3332. .15787)
/gene="MT2108"
complement (3332. .15787)
/gene="MT2108"
/note="similar to PID:987099; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polyketide synthase"

```

FEATURES
    source
        Location/Qualifiers
            1..43280
                /organism="Streptomyces fradiae"
                /db_xref="taxon:1906"
            1..43280
                /gene="tylG"
                816..14234
                /gene="tylG"
                /note="ORF1; putative"
                /codon_start=1
                /transl_table=1
                /product="tylactone synthase starter module and modules 1
                & 2"
                /protein_id="AAB66504.1"
                /db_xref="GI:2317860"
                /translation="MSSALRRVAVQNSCGVGLDMLTNTAAQNTGDQEDVDGPDSTHGGE
                IAAVTDQHRMLELGEWELLSRGRMPTQDDGTWRAALEDHAGFDAGFFGNAR
                QVATDQHRMLELGEWELLSRGRMPTQDDGTWRAALEDHAGFDAGFFGNAR
                GYTATGLHRAALNLSHFLGRLGSLVVDSSQASLVAVQLACSLRGETSLAVG
                GYNLILTESTTMMRMGALS PDGRCHTDPDANGSVRGEAGVVLKPLDAAADG
                RYCVIKGGAVNVDGGSLSLTPDRAQEAVALRQAVRRAGVSTGAIRVVELHGTGTRA
                GPVRAALGAVLGAAGDSRSTPLRAQEAVALRQAVRRAGVSTGAIRVVELHGTGTRA
                VPSLFTNPDI PLDDLRVQTERQEWNEEDDRPRVAGVSSFGMGTNNVHLVIAEA
                PAAAGSGAGGAGGAGISAVSGVVPVVSGRSVVVRREAAGRLAEVVEAGVGGLA
                DVATMDRSRGYRAVLARGAEALGRALRAGDGDAGVVTGAVLDGCVVVGGLA
                CGAAGAGAGAGAGGAGGVLVFPQGTQWGMGALLGSSSEVAASRCARALSV
                HVGMDLLEVSGAGLERVDVQVPTWAVMSLARVQAMGVDAVAVVGHSGQETAAA
                TVAGALSLEDAANVALRAGLIGRLVLAGRGANAAPLPAGEVEAGLAKWPGVEVAAN
                GPASTVSGDRRAVAGTVACQAGVQARLIPVDYASHSHVEDLKEGLERVLISGIRP
                RSPRVSTVAGEQGPVFDAGVFNRLNRVFSVAVGGLLEGGHRRFIEVSAHP
                VLHAIETQAEADRSVHATGTLRRQDQSPHRLTSTAFAWAHAGATLTWDPALPPCHL
                TLTPTVFNHHYMLDITDGGQDDATQKSGPLTRELRLGULPSSOKOLGFLDLVCR
                HTAVLGLDTAAEVDPDLSFKKQISQMTGVELRNLLTETGLALPTTLVDRPTPRA
                LAQPLHLLDSCSPGSVLA PAQKSEAEQETAVVMGCRPGCGVSPBALRVLVEG
                VDASFFPDGRDVEGLYDPDPGAGKSVREGGLDHADEFDAEFFISPREAVAM
                DPQRLLETSWEALIRAGIDPHSLHSGRTSYAGVPOEYGPRLAEGAGSDGYLLT
                GTSGSVGRVATYTLLEGPAVTVDTACSSILVALHVAQALRGECOMALAGVTVM
                AGCFMFPBSRGGLAARDCKAFADGADTAWAGCAVVLVERLSIDARLGHVPLAV
                VCGSAVNQDASNGLTAPSGPSOERVIQALGNARLTAVADVVEAHGTGRLGPDIE
                AORLLGTGRRDGRPPWMLSLKSNIGHAQAAAGVAGVIMVLAARMYGMPLRTHLD
                EFSRVDMSAGVWLLTEAREMFPVDRPRAAVSPGVSGTNAHLILEAPDTEAESB
                TTPVRSSEVSAALDVGVPVVSGRSVVVRREAAGRLAEVVEAGVGLDAVAVT
                MAGRSRGYRAVLARGAEALGRALRAGDGDAGVVTGAVLDGCVVVGGLVFP
                PQGTQWGMGAGLLGSSVFAASRCARALSVHVDLLEVSGGAGLERVDVQV
                VTMVMSLARVQAMGVDAVAVVGHSGQETAAA TVAGALSLEDAANVALRAGLIG
                YLAGRGANAAPLPAGEVEAGLAKWPGVEVAANVNGPASTVSGDRRAVAGTVACQ
                YLQARMLIPVDYASHSHVEDLKEGLERVLISGIRPSPRVSTVAGEQGPVFDAG
                YFNRLNRVFSVAVGGLLEGGHRRFIEVSAHPVLHAIETQAEADRSVHATGTLR
                RQDQSPHRLTSTAFAWAHAGATLTWDPALPPCHLTLTPTVFNHHYMLDITPTTPT
                TTQSPPTDAMRYVTWKALTESSPVRPHSIGRCLLVAPPTTGDGLDGLTVLISERGA
                VARLEVPICARAEVAELKPSMESAGEBENTTVSLGLVLPSTDAVTSIALQOAVSD
                IGVPAARVWALTRAVAVVPGTTPDAGAQWGFGRVALELPDINGGLIDLPETAEL
                TITPETSQPTPERLPUPNRRALLEAAVLAGRDGEDQVAVRAGSPGYGRRSRAAA
                AGAASQWQSVTLITGGMGAIGRRLARRLAAEGERLVLTSRRGPEAPGAELABELR
                GHGCEVVAACHDAVDAALVATYPPNVAFTAGILDADVIDLSPESFETVPTGAK
                VCCAELLHQTADI KGLDAPFLVSSVTGTWAGOCQAVAAANAALDALARERRAAGLP
                ATSVANGLAGGSMAGAGEESLSRRLRAMDPDRAVDALLGAMGENDVCVTVVDVM
                ERPAPATNAIRGRCLPDTPEAREALTAAGTTSATPDGAPELARLSMLNTERLKF
                VELVTEAFVUNHPNTDAIGERPFKSGFDSLTSLEUNRLNAGTGUKLPATVIFD
                HSPSTALRLDLRLTGAGAPAPADEPPLVAVADDDPVIVGMACRPPGAGVTEA
                LKWLVTEDRVDVGAAPTDRGWDLSYDPEPGVAGKYVREGGLHDAEPFAEFGI
                SPREAVAMPDQRLLETSWEALIRAGIDPHSLHSGRTSYAGVPOEYGPRLAEGAG
                EYEGYLLTKCSASVSGRISYTLI KGLDAPFLVSSVTGTWAGOCQAVAAANAALD
                LAGVTVMAAPGLVFEFSRQGLAADRCFAFDGADGTAWAEGGVVLVERLSRAR
                LGHPVLAVCGSANGQDASNGLTAPSGPSOERVIQALGNARLTAVADVVEAHGTG
                TRGDPIEAALIGTYTRDRDAECPVWGLSKSNIGHAQAAAGVAGVIMVLAARMYGM
                LPTLHVDPEFSRVDMSAGVWLLTEAREMFPVDRPRAAVSPGVSGTNAHLILEAP
                EALEALEADPAEAPAPADPTDVTEALEAPDAEAGAKAPGSPPEAPQAVGVV
                VVSGRSRVVVRREAAGRLAEVVEAGVGLADVAVTMAGRSRGYRAVLARGAEALAG
                BURLAGGDPDAGVTVACQAGVQARLIPVDYASHSHVEDLKEGLERVLISGIRP
                SHRECARALSVHVDLLEVSGGAGLERVDVQVPTWAVMSLARVQAMGVDAVAVV
                VGHSGQETAAA TVAGALSLEDAANVALRAGLIGRLVLAGRGANAAPLPAGEVEAGL
                KWPVGQVAAVNGPASTVSGDRRAVAGTVACQAGVQARLIPVDYASHSHVEDLKG

```

```

ELRVLSGIRPRSPRVPCVTVAGEQGPVFDAGVFNRLNRVFSVAVGGLLEGGHRRFIEVSAHPVLHAIETQAEADRSVHATGTLRRQDQSPHRLTSTAFAWAHAGATLTWDPALPPCHLTLTPTVFNHHYMLDITDGGQDDATQKSGPLTRELRLGULPSSOKOLGFLDLVCRHTAVLGLDTAAEVDPDLSFKKQISQMTGVELRNLLTETGLALPTTLVDRPTPRA LAQPLHLLDSCSPGSVLA PAQKSEAEQETAVVMGCRPGCGVSPBALRVLVEG VDASFFPDGRDVEGLYDPDPGAGKSVREGGLDHADEFDAEFFISPREAVAM DPQRLLETSWEALIRAGIDPHSLHSGRTSYAGVPOEYGPRLAEGAGSDGYLLT GTSGSVGRVATYTLLEGPAVTVDTACSSILVALHVAQALRGECOMALAGVTVM AGCFMFPBSRGGLAARDCKAFADGADTAWAGCAVVLVERLSIDARLGHVPLAV VCGSAVNQDASNGLTAPSGPSOERVIQALGNARLTAVADVVEAHGTGRLGPDIE AORLLGTGRRDGRPPWMLSLKSNIGHAQAAAGVAGVIMVLAARMYGMPLRTHLD EFSRVDMSAGVWLLTEAREMFPVDRPRAAVSPGVSGTNAHLILEAPDTEAESB TTPVRSSEVSAALDVGVPVVSGRSVVVRREAAGRLAEVVEAGVGLDAVAVT MAGRSRGYRAVLARGAEALGRALRAGDGDAGVVTGAVLDGCVVVGGLVFP PQGTQWGMGAGLLGSSVFAASRCARALSVHVDLLEVSGGAGLERVDVQV VTMVMSLARVQAMGVDAVAVVGHSGQETAAA TVAGALSLEDAANVALRAGLIG YLAGRGANAAPLPAGEVEAGLAKWPGVEVAANVNGPASTVSGDRRAVAGTVACQ YLQARMLIPVDYASHSHVEDLKEGLERVLISGIRPSPRVSTVAGEQGPVFDAG YFNRLNRVFSVAVGGLLEGGHRRFIEVSAHPVLHAIETQAEADRSVHATGTLR RQDQSPHRLTSTAFAWAHAGATLTWDPALPPCHLTLTPTVFNHHYMLDITPTTPT TTQSPPTDAMRYVTWKALTESSPVRPHSIGRCLLVAPPTTGDGLDGLTVLISERGA 1" /function="encodes acyl carrier protein domain of module 1" 8541..11370 /gene="tylG" /function="encodes module 2" 8541..9812 /gene="tylG" /function="encodes beta-ketoacyl ACP synthase of module 2" 10260..11246 /gene="tylG" /function="encodes acyltransferase domain of module 2" 11319..11876 /gene="tylG" /function="encodes dehydratase domain of module 2" 12861..13415 /gene="tylG" /function="encodes beta-ketoreductase domain of module 2" 13719..113970 /gene="tylG" /function="encodes acyl carrier protein domain of module 2" 14351..19945 /note="ORF2"

```


Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
 Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites
 Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
 21477403
 2 (bases 1 to 78210)
 Ikeda, H.
 Direct Submission
 Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan
 (E-mail: ikeda@c.kit.ac.jp, Tel.: +81-3-5791-6242, Fax: +81-3-3444-6197)
 Location/Qualifiers
 1..78210
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 /note="Actinobacteridae"
 complement(1..1254)
 /codon_start=1
 /transl_table=1
 /product="putative lipoprotein"
 /protein_id="BAB69301.1"
 /db_xref="GI:15824137"
 /translation="MGEALITLCTSTRLLAATAGVLGGVLAVALTALGTSNAATSGNAGHPTSOARINOAQAEASDARIETPGOGTYSGTNPVSVTVSGKTKVMTTAVATCAEIPGLTSTGTSKPNGLERATKYQIAABAGDVKGRSTTGATITVSPANDFI GHLSPEDGTVVAMPVNFDFKADKAAVSEIRVSSSGQVRVWFNDRHVRDIFRPNYRPGSTVTVLNHGIQKTVFKIGRSQISTDAKTQMTVVRGDKTKTIPIS SGSPHPTVGMQVISEKFRQIHMGASVGLTKNGKPSYDIKAVPHAMRLTDSGTFI HCNVGTDFVFCBEANTSHGCVGLKDVGRGSHGRQPAAFEDHSIVGVDVIVKNSSEKT MLAPNGLSDMNMSSEWAGSPAT"
 complement(1256..2290)
 /codon_start=1
 /transl_table=1
 /product="putative transposase"
 /protein_id="BAB69302.1"
 /db_xref="GI:15824138"
 /translation="MPVEELSHDQVSRVGRFASAPTPGEEQFRLDLAKALESARKR ANRLGWAQVQVQVRLGVPFTEDELSVVEVVRFAAEQGLVDGFEAYGAKQNR YBHEIRDAYAYRSPAAEQVREFLTVWASWEGPEGLDFRAVGLWVNRVLLPG IITLLRAVEAQAENALYRTLDEAPKLRQSMRDLKVPGRVRSLELRSPM GYVGSADRLAFSEIRMEAVMTAGPSTDIKELRVSVQSVQRWSGLADRGPRG AVQPVSVSELTFLPAALDEERAKGPASMTIRYLHDDHPFIFQRYAYKWSQGLHSP GRSAQYRQ"
 2935..21081
 /gene="pteA1"
 2935..21081
 /gene="pteA1"
 /codon_start=1
 /transl_table=1
 /product="modular polyketide synthase"
 /protein_id="BAB69303.1"
 /db_xref="GI:15824139"
 /translation="MVESLNATGNRDCSASADGIAVIGLACRLPGAADPAAFWR LLSGADATLVPVDRWDGAVADASAPDRTGIRRGFLDLSGDGDFGLSPKEAA ANDPOORLVLELWALEDAHVRAATLRSTRGVFGVGMDDYATLHRSGLTALSPH TVGLHRSIIRNVSTYGLNGPSTLVDSQSSLSVSHLACESLRKSSSTIALAGV NUNIVESALNKAAGFGLSPDRCFTFDARANGYVRGEGGIVWLKPLARVADGDP YCIVIRGSVNNDDGDLTVPLQSGQEVRLAYRAGVNPADVGYVELHGTGTVKVO PTEALGAVALGTGRPRPLRVGSATNVHLEGAAGITGLLKAVLSLHRLPASL NECTNPATPLDRLNLRVOTESHSEWAEDGRPLLAGVSFGMGNTNCHIVLASGAVE AAGEGATAGGAAPTFWPLSAKTASGLRAQAALLAATEAPLGLAPDVGCWLSATGRTA FERRAVVGEVRDFFDLRLKLELSTGGIDAAITLGRAGRGELAFVFGGSGRAGMR ELAESVPFASALDECAHLDPLPRLEVPMPAPVGTGEAAELDQTLVTSFLAVE VALFRLESDWITPDVIMGHSIGELAAHVAGVLSADACALVAARGALQALPAGGA MIAVQALEDVEARIGERTDRSIAALNGPDSVYSGDEDLATEADAPALAKRTSR LKVSFAVSPHMEPLABRNVAEGITFHAIPRIPIVSNVTYRGLSEPSGECYTAAYW VZHVREAVPADGVARLDEQGVTYLELPGQVILASMARAGDSLSFVPALTRRPE AQALLASLHVHGVPEPDMDALFGGGRGARRKVALPTAFERQRLWDGDTTSPAT TGWAPVAPQPDIAETDSAPVGLRGLAERSEDERDEAVLDWVRAHTAAVLGHAEAR QVETEWTFKDLGFDLSLSSVELNRQLSEATGLRLPSGLLPHDPTPAALANHLAAQSLGG

GDAAAAATPATVADPHEPIAIVANSCRIPGGVSSPEDLWRLIASGGDGIISGFPTDRGMD
 VETLYDPEPGTSGKSHGGLHEAAEPFDAGPFGIGPREAAAIDPOORLVLETAWEA
 FERIDPATLRGSRAGVFIGATADQDQPRLEHEDPADGVEGVLTTGTTTSSVASGRVAYS
 FGLGPAVTVDTACSSSLVALHLAVQSLRSSECTMALAGVTVMTSGMPFVPSRQGR
 LSAQRCKFAAGADGTGAEGVGMGLVERLSDARRNGHGVLAIVRGGAIVNQDGSNG
 LTAPNGPSQORVRAALAAAGLADPADVAEHGTGTLGDPFIEABALIAVAGQDGRG
 DRPLWGLSKNSIIGHTQAAAGVAGIKVMAMRHVLPRTLPHVDEPTPHVDMWSASGVR
 LLTEAVEPARDLPRAAVSSFGISGTTNAHVIVQNAATSEBPARGTAVAGVPVWVSA
 KSDVALROAGRLSFLDAEGVGGVVRPDEVALSLATSRAVLERRAAVAVGCDLAEAL
 KLEALASGASGAAPVGGVGRVFLPSGOGSORIGMGRRELYAIVFAAAYDEVCA
 FLDPAPVDVSEELNGTGTTPALFAVEVALFRLESIGRIDYVAGHSVGEIAAHVA
 VLSLEDAALVISAALMOALPDGGMVAVOATEDEVLRHLTDGVGAIVAGNPKRSIP
 VSGAEDAVTAGEIFAGRKTSLKIVSHAFSLPMDPMLPEFAAVVGLVTFSSRIP
 VVNSLTRGAEVTVYVVRHREAVFADGIEITLGEIVTVFVGGVGLSALAQF
 CFDDGTDIVTVLRADEPEVALVTAIGRLTHGISPDQVLPFGARRVLPVTAFO
 RERYLDAARAATAGADAMDAEWDVSESDRASLGLLELTDVLSAKLSAWR
 KRRRQSTTDGWRIRITWOPUDDMVRAAPSCTWLYVYSEETAWEAIRAGLTGELVAL
 VPFVAGEDTDRDALAPADAGQBOAGILFAAQGEAGANALQVLLVLAQAGDAGI
 EAPLWCLTSGAVSTGTDADLTPVSAQLWGLRIVALEPQORWGLVDLPABLSRAL
 ERLAGLLAQADEDLAVRASGVYGRRLVRAPLSAAPADAPMSPRGTVLVTGCTGALG
 QVRLWAGAGAEHLVLSRRGPDAPGAELKAELEELGAQVTVVACDAADDAELALF
 GRYLVNAVHTAGVLDGLVESLTPERLNVLRPKVDAALHLHETDRQDLDAVLP
 SSMTGWNGCGGAYGAANAFDLAALAEHRRAGQVVALAVANGSWADGCMAGADHL
 RRRGVRAIALPAISVLHGALTGETSMTADVDMERFVPAFVTPCPLLLGVPPEAR
 EALBARAAQSPVPSVALVQRLGATPGQGRILLDLVREQATVLTGHTKGFAEA
 DRTFRETGFDGLTAVELRHLNLTATGLRLPTTLVPHDPTLARHLEELGLRHGLE
 TAAQATAAAADAPLTAIVGMSCLPGCVQSPEDLWELLSSRRDAMSFPDTRGDWID
 ALYDPDRPKTYSREGGFIQADRFDAFLGICISPREALANDPOORLLELTAWAPE
 RAGLAVASRASRTGVFIGTNGQDYANGLNAPEIEGYALTGKAASVSGRISTYFG
 LEGPACTVITACSSSLVALHLAQAALSGCSMALVGGVTMTTDPLEFVSRRGLS
 PDGRCKFAAGADGTGEGVGLLVERLSDARRNGHGVLAIVRGSAVNOQDGSNGLT
 APNGSQRVITROALASAGLTADVDAEAHGTGTRLDGPIEAQALLATYGOEHSDDR
 PMLGSKVNSIIGHTQAAAGVAGIKVLMAMRHVLPOTLHVDEPTPHVDMWSAGVRL
 TEPVQWPGTTPRAAVSSFGISGTTNAHVIVQNAATSEBPARGTAVAGVPVWVLSA
 SEVALAQRLELAFATDADDVDFDAGFSSATITRERALEHRAVIGSDPAELRAGLEBALA
 AGEPAANVYAGRAHSADKVGFLPSGQSQRGLMGRELYAIVFAAAYDEVCAVLLD
 VDVAETLHOTGTQTPALFAVEVALFRLESIGRIDYVAGHSVGEIAAHVAGDGL
 DAAKLVSARAAKLAKPAGVMAVQAADDEMLPHLDLAGIAAINGRPSVIVSGAE
 DAVTAIARVPAQAQRKTSRLKVSHPHSLPMDPMLPEFAEVVRGTVPEPVIPTVSNL
 TGRLTESYTPYVVRHREAVFADGVLTLHELGVTTVEIGPGVLSALVQGLCDDG
 LDIITIPALRADVRPEQALTYAQLHVGAEVNDHAFPGCARRVVELTYAFQERERY
 TDAFADAGDMRAVQGEAGHFLGAAVPLADSGLHLTGRLSKTHPLWLVHVASGVA
 LPLGTAFVELAITAADAVGCDLLELTETLTPERGVGVALQVHVGADDSGRPLF
 VHSOITDGDVAHEDESERVHASGVLTATGTPGVDSGLTAMPVGAEPVVDGVPY
 DRLAGMALDYGVPQGLRAAWRSGDDPFAEVELRCABRTDAGVGLHPALDPAALHTV
 VLGAEPEAGTGNGLLPANSGVRLAAGASDLRVKVSRACTGTSTVSLVLDGTTGGP
 VAEDVSLTRPVPDQLRGSDDDAFLKLVPSLPSPADGPKRIETEDLALURAT
 DDPADTEADAAAAADAVVPCPTGTGADLATRVRETHAVLELVQWLAERPGRLVL
 VTRGDLAQSAVCGLVSAQTPENPDRIVLVETSSDAHAGAGSGVDDADLARVPLGLIA
 SGEPOAAVHADDEVVRPRLVRVAVPDAEPGLGSGTLLTGTAGSLGGLPARHLVA
 EHGVSLLLVSRGSDAPGAELTADLTAAQADVTMAACDVADRAAVALLAGPCQSL
 SAIITHAGVLDGGLIGSLTPERLDANFPKVDALNHLHLAGELSGDLSPVLFSSVA
 GTLPGQNTAANAATFLDAEHRRAAQRATSLANGWLMAQDSSSMTGSLDHTDLT
 RIQRMGAAIIPSSDGLFLDAAALDRAVAPLRLDGLGQTPVPLURGLRGVTRGT
 ARRTAQAPKSLGORLAGLAEREQVLDLVRTEVAALVGHAGAQSAGVADDSFPDI
 GFDLSLTAPEGLNRLNSAVGMRLPATLIPDYENPLTLAHLTGEAAGGAATAVPTLS
 TATCTADSPIAIVGMACHYPCGVQSPEDLMOLVFSGRDASVSGPGRGDMVEOLYDPN
 PDWGTSTYRGGFLHDAADDAEPFGISPREALANDPOORLLELTSEWEAREGADIP
 FQMGTSRTGVAGVWYHDYGRVHTSPGLEGLVNGSAGISASGRVSTYTGLEGPV
 TVDTACSSSLVALHLAQAALSGCSMALVGGVTVMAGSVFVPSRQGRSLADGRCK
 AFGAGADGTGAEGVGLLVERLSDARRNGHGVLAIVRGTAIVNOQDGSNGLTAPNGS
 QORVROALANAGLSDESDVAEAHGTGTRLDGPIEAQALLATYGOERSEDRPLWGS
 LKSNHSQAAAGVGVTKVMAMQHGLPOTLHVDEPTPHVDMWSAGVALLTEAVEM
 PASDGPRAAVSSFGVSGTNAHVIVQNAATSEVTARCTAVAGVPVWVLSAKLSAALRE
 QAGRLSLDAEGVGGVGRPEDVAISLATSRAVLECAAAVVEBELGFRQGLEGISV
 GAAPVGGVAGRGVGLFSGGQSQRIGMRELYAIVFAAAYDEVCAVLLDVTVDVAE
 LTHAGTQPALFAVEVALFRLESIGRIDYVAGHSVGEIAAHVAGVLSLEDAKL
 VSAARLMOALPAGGMVAIQAETAEVPLHTHGVGLAAINGRPSVLSGSDVLDI
 AEVFAQQRKTSRLKVSHPHSLPMDPMLPEFAEVVRGTVPEPVIPTVSNLTGLLAE
 VYIPEYVVRHREAVFADGVLTLHELGVTTVEIGPGVLSALVQGLCDDGTHLVTA
 PALRADPEPVLVAVGQLHTHGISPEWQALFPGARRVDLPTAFQERWLDAPQA
 SGDVRAGLGAADHPHLLGAAIATADSDGVALTGLSLGTHPLHADHVAIVLLPGTA
 FVELAIRAGDETGCOWIBDLTLELPLVVPVPERGVTLOVAVGPEDESGRRQLSVHSRVD

QUALVLGHATGGDNDAGRAFRDGLDFSLTAVELNRLNNTVTGLRLPATLVFDYPTVRH
LATYVLELLGTDAEVATQPAAVADDDPIVIVGMACRYPGCVSSPEDLMRVLTEGT
DAVSGPTNRGWDVSLYHPDPDHPDPTGSTRGSGFLHEAGFDPGFGFSGPREALATO
SOORLLLESSWEALBRAGIDPVSRLSRSTGVAGVWVSYSAMLASPEFGFGSGSS
PLASGRVATYGLLEGPAVPTACSSSLVAMWAMQALRSCEGCLALAGGYTVMSTP
AVDFARQYGLPDRCKAFDAADGVGSEGVVLRLERSDADVARNHETILAVRG
SANNQDSNGLITAPNGSQQRVITQALASGLTAGDVVDEAHGTGTVDGPIEAQA
LLATYGRDEPREPILLLGSKNLGHTQAAAGVAGVTKMLAMRHGVVPTLTHVDAPS
SHVDSGAVELLSEQAAPETGRVRRAGSGFSGTNNVIVBOPAKAIAAAGA
ARTPCAVLLSGRSALRGQAPALGHLOAPDAELVDVALSLATRSFEFORAA
VVAQDQOLIASIGALAAADRPDPVAVEGAAGRTAVLFTQCGSORAMGRELHEVQ
PEFAAFDAVAFDPLDRLPREVFAEDSGDEAALDETGTQPALFAVAVLFRQ
VESWGRVDFAGHSGETAARVAGVLTEDACRLVAARATIMQALPTGGMAMIAFL
TEDEAAHLDDTVAAAVNGPQSVISGDEEAETIAATFAERGRKTRKLRSHAFHS
PMDGMLDRTVAAGLTYRAPRIPLVSLDTRRADDAEVCATYEVVRHVRRAVPAD
CVTLRDACATYFLGSDGLLTAMAEDTLGDHDAELVPMLRAGRAELAAATLAR
LVVRGVDVMAAYLAGTARRDLDTPYAFQHAYYWPQLTPPAALAAADPADQOLWAA
VERGDAREALILGEGQDLTPDLSLLPALTSWRNGEKHLLDTLRVDETRUSKP
TAPVLDGTWLLVASDAAQDQALLDADALGSHGARVRRLLDDSCADRAVTAERL
ARTADVDAATVLSVLPDERDADDCPPLTRGLALTVALQALADTGAQRGLWATRG
AVSTNPADPVTHPVQAAAGLGRVLEHPLRWGLVLDLPQVDFDRAGORLAGILAVK
DAPDGEDVALRATVSGRRLVRHTVEALPTAAETATGTVLTITGSGTGLGAEVARNL
ARAGAHLVLTSEGEDPAPAAELRAELREGYSPSVACDVADRDLAAVLTALPEE
LEPTGVRFAGVGHGPDLTSTAEFAGLTAALAGAAHLDALLADRELDFFVLFGRS
AGVWGSGNSAYGAANAYLDLALHRRARGLAATSVMGPAEAGMADDAVSETLIR
OGLGLDPAAPAMTELRRVAVRQDVTTVADVDWQRYAPLFTSARPSALITAGLEVRAL
AADERTQDAGSEVTVRVALABEQLRLTLDTVRTEATVILGSHSADAVPEGRAF
RDVGFDLSLAVELRKLGAATGLSLPSWVFDYPTLELAOYLRAELIIGLVEAGPV
ATGADDEFIALGHACRFPFGVSSPEQLMDLVAGTDALISEFPVNRGWTQHLFPDP
PDRGTYTSGGHEADFDPTFGISPREALVMDPOORLLLETTEVNSFERAGIRP
ETRLSTLTGTSSYSGVGLGAGDTEGHMTGSSPVLGRLSYFVLEGPATVTD
TACSSLSLHALACSLRNGENLAVAGGATIMTTPNPFIAFSORALAKAGCRKAFS
DADGWTALAEVGVVLVERLSAORNGHPLVAVLRGSAINODGASNGITAPNGPSQVS
VTRQALAAAPGDIDALEAGTGTPLGDIPEAOLPATYCDRDPESALGSGKRS
NIGHTQSGAGIASVIMWMLRHSELPTLHADASSHVDSAGTVRLTLQARWPEP
GHPRAAERLISGNTAHVLEQAPVADPAEERPAVAPVPIAGVVPVWVPTAESAA
ALRQAERLLIAGHTVTPAAGPLDGLSLVSARAFPEHRAVVPVAGTDFPLAAR
AVATDPSVAVRAGVADVGVTFVFGOGSOWMGVOLLDESFAVFAERIAECAAL
AFTDWSLVVLVRGVGAPSLERVDVQGPASFAVMVSLAALWRSRGLPDAVVGHSQ
ETAAVNGSALSIRDOGARVVALRSQAI GHALAGRGMSVALSDVLEPRLVEFGRV
SVAANGPSVVVAGEPEALDHALHATDADRARIADVYASHSHQVEDLHEELLEV
LAELAPRTSEVFFSTGTDWLDARTMDAGYFNRNLGRVRFADAVDALLAAEYAFV
EYVSHPLVMSVQEAIDEAGVPAVAGTLRRDQGGTDFLLSAAAEVFGVSDVDWAGL
FEGTGASRIDLPTAFOHLEHWAPEVPAEVAADDDAAFWTAVEDGSDVSALTALG
TDEDSVAAVLPLATSWRRARRDRSTVDARVYVANKPLGTLPHSLTGTWLLVADG
IDDTVAGALETYGAEVRLVLDSECDRAVLRLAGAEVDVTGIVSVLAAERTDAV
PGLSLVGTALVALIQALGDAEIDAPVWALTRGAVSTGRADDELTAPOAQVTGLGW
AALEHPQRWGGTLDLPALDARAQRLAALVLSGLSDDDQLAIRPSGVPTFRVRAEA

Query Match 54.1%; Score 133; DB 1; Length 123580;
Best Local Similarity 71.4%; Pred. No. 1.5e-08;
Matches 175; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1 CTGGCGCGGACACACAGACACACACACACACACACCTCTCGCCCTGCTCGCTCCAC 60
DB 29330 CTTCGGCGAGCAGCCCGCGGACACACGACGACCTCTGCTGCACTGCTACCGAGAAG 29389

QY 61 ATCGCCACCGTCTGGGCGACACACACACACACACCTCCCGCGACCGCGGTTCCGC 120
DB 29390 GCGCGAGACGCTCTCGGCGACCGCGGACGACCTGCTCGAACCCGACCTGCGCTTCGC 29449

QY 121 GACCTCGGCTTGCATCTCCCTACCGCGCTCGAAGTACGCAACCGGCTCTCCCGCACAC 180
DB 29450 GACCTGGGCTTGCATCTGCTGACCGCGGCTCGACCTGCGCAACGAGCTACCGCGGAAC 29509

QY 181 GGAATCCGCTCCACACACCTCGCTTCGACCAACCCCAACCCCAACACCTTCACCCAC 240
DB 29510 GGCCTCACCTCGCGGACACCTCTGCTTCGACCAACCCCAACCCCAACCCCGCGCC 29569

QY 241 CACCT 245
DB 29570 CACCT 29574

RESULT 11
AX211739
LOCUS 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGE TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source 1. 125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN
Query Match 54.1%; Score 133; DB 6; Length 125401;
Best Local Similarity 71.4%; Pred. No. 1.5e-08;
Matches 175; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1 CTGGCGCGGACACACAGACACACACACACACACCTCTCGCCCTGCTCGCTCCAC 60
DB 29738 CTTCGGCGAGCAGCCCGCGGACACACGACGACCTCTGCTGCACTGCTACCGAGAAG 29797

QY 61 ATCGCCACCGTCTGGGCGACACACACACACACACCTCCCGCGACCGCGGTTCCGC 120
DB 29798 GCGCGAGCAGCTCTCGGCGACCGCGGACACGACCTGCTCGAACCCGACCTGCGCTTCGC 29857

QY 121 GACCTCGGCTTGCATCTCCCTACCGCGCTCGAAGTACGCAACCGGCTCTCCCGCACCCAC 180
DB 29858 GACCTGGGCTTGCATCTCGCTGACCGCGCTGCGCAACGACCTCACCGCGGAAC 29917

QY 181 GACCTTCGCGCTCCCGCACACACCTCTCGCTTCGACACACCCCAACCCCAACCTTCACCCAC 240
DB 29918 GGCCTCACCTCGCGCGACACCTCTGCTTCGACACACCCCAACCCCGCGGCTTCGCGCC 29977

QY 241 CACCT 245
DB 29978 CACCT 29982

RESULT 12
AB070949
LOCUS 78210 bp DNA linear BCT 11-OCT-2001
DEFINITION Streptomyces avermitilis polyene macrolide biosynthetic gene
cluster.
ACCESSION AB070949
VERSION AB070949.1 GI:15824136
KEYWORDS
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,


```

/gene="pimS4"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS4 protein"
/protein_id="CAC20919.1"
/db_xref="GI:12055070"
/translation="MFNDKPNKKLVYLYKWNVDLHQTQRRLREVESAKQBPFIAG
MSCLRGVGVVPEFQWLLKDGKDGISTPTPTDGRWDLDDLAGSGSSATVGGTSDID
ASPDAGFGVGSIPREAVTMDPQOORILLSSWEALERAGIDIPVSGTGTATVGGTGD
YINTVMSREDIEHGSGTGLIASILSGRVSYTFLGEPVAVTVDTCSSLSVSLHAAQ
ALRVGCSLALAGGVTAISTPMSFVGRSQGLAPGRCFAFSDTADGTANSEGVEVL
VERLSDARRNGHVLAIVRSVAVNQDGSNGLTAPSGSQQRVIRQAALAAAGLASD
VDVVEAGTGTALGDPIEAQALATYDRDRDPERPLLLGSVKSNGHTQAAAGAGLI
KMLAMRHGVLPRTLHVEAPSSHVDWSSGALOLLTPOQWPKADRWAGVGSFAISG
TNAHIILEOPPAAREDEDEGHTSPSAVPLVSAKSEALTAQLDRIISALDRSPLD
LGLSLATTTRAHLEHRAVLISEGGETVEVARGHNEHTAGKLAVLFSQGGQARVGMGREL
YGRFPVAFALDGVLAHFDAGLURDMFGAEGLDETGTQPALFAIEVALFRLSBSL
YRDFVAGHSIIEIAAHVAGVPSLADACALVAARLMQALPAGAMVAQAAESEV
AARLVDSVIAAVNGPEAVVIAEEEREVLRIAEFTGLGRKTRRLPVSHAFSPMEP
MLEDFRVAEGLSYEAPRIALVSNLTGELATDVLRSAYVWRHRETVRPADGVASG
AAGASVFLGLGDPVLTAMAOHTLDGTATVPALRKDRSEETALLTALAQLHVGVC
VWMSGIFAGTGARRVDLPYPPQHERYWRPVAVSGDVVAGLRSNEHPLLGATLILA
EGEVVITGRLSLQATPMLTDMTSGSVLPFPATPFLLEARNADQGCCDTRDVEFLMT
PLVLAESGTQVVLVPADAEAGTRTISIHARPDPDPDPQWQHAGGTLTSGERIVDF
GSTVWPQGAEEAVDLLEGFDYDTGYPAFQGLRSVMQCDGAYVEVALPQVADAGAF
GLHPVLLDAVLOSRRMAGIAGAIDTFLPFAWGVSLHAGCASVRLARVITKEDTSL
VADAGAPVLSVEELVLTOLTHTDAPLTARRGAERDALLSLEWVSGPEVPAGVRC
VSLDALGLVGSATLADLSDGDEFVVPVSGVGDVPAHAHELITARTLIGLQWBSG
RGTSASRLVFTRNAVASDGSVDAVAAAGLVRSAQSESPGRTIALVLEADSG
DLAEVLGALLGLAADAQFVVRREGAVRVGRVSENEGASVFEWDADGTVLVGGT
GLGLRELAHLVVERGVGHLLVSRSPGASGVGLRDLAAGHAEIAVRACDVADRA
AVEAVLGAADAAHPLTAVHTHAGVLDGVIASLTPELRSVLRPKVDAANWHEATKD
LDLAAVLYSSISVGLSGQNYAGNVFLDALAGHAGLGPAGQSLANGWATOGGK
MTATLSDAMQWASGTAPLSPERGLALFDAAGTAGSATHIIVAGRVSPVRAQGPVP
SMUGLVGARRTAVNTTGTAVSLADRLURARPEERTQMDTMDVTEAAALVGHASS
QNDARREYFELGSLTSVELNRSLTGTGLRSLTAVTSLVTVFVDSKNTPLDAARVILAA
QALDGSQAQGLPVAAPETDSLERMFLDALDSKVPQAORMLSALGRVSHFNTA
ELEDPLTAEAGCTPRLI CVSTPTANGVHEYARLAASFRGRHVSALPLVGAT
GRLPATAVRVRIEASALRAGNPPVVLVGHSSAGAFYLAALLENWTCIRPEAV
VLLTSLLRHONESIDYAGLRMRHVMDEVPVRNTRLSAMARWGMMLNQLVHRV
TTPVLIIRATKETFIGITGTIGYEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
WLDLSGNA"
complement(9772..15198)
/gene="pimS3"
complement(9772..15198)
/gene="pimS3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/translation="MTTSSETVVAALRASLKEAENLRQNRKLVAATEPIAIVAMAC
VGGVQSPSEWELLTAGKDAIGAPPTDRGWDLDALRAGVDREGNVSRQGGFLDG
YGFDDAFGFIIPRESVMTQOORILLTSSWEAIERAGIDPTTLGRSGRTVGITNGQ
DYAVLLVSLADATDGTGIAASATSGRLSYTLGLEGPVAVTVDTCSSLSVALHTAS
HALRAGCSLALVGVWMSVSPGSLMFSRQGLAADGRCKAFSDTADGTGWSGVGV
LVVERLSDARRNGHVLAIVRSVAVNQDGSNGLTAPSGSQQRVIRQAALAAAGLAS
VDVVEAGTGTALGDPIEAQALATYQGDPRFPLRLIGSVKSNGHTQAAAGAGLI
IKMLAMRHGVLPRTLHVEAPSSHVDWSSGALOLLTPOQWPKADRWAGVGSFAIS
GTNAHVI VQAEAPDEPVAEPGLAPVAPVPLVSAKSEALTAQLDRIISALDRSPLD
VGFSLVAGSFEHRAVLLAQCPDGLSEARGRAETSRPLAVLFSQGGQARVGMGREL
LVVRPFAEALDLEDAHAGLURDMFGAEGLDETGTQPALFAIEVALFRLESIL
GVRPDPVGHSGIETIAAHVAGVPSLADACALVAARLMQALPAGAMVAQAAESE
VAARLVDSVIAAVNGPEAVVIAEEEREVLRIAEFTGLGRKTRRLPVSHAFSPIME
PMLEDFRVAEGLSYEAPRIALVSNLTGELATDVLRSAYVWRHRETVRPADGVAS
LEAGASFLGLGDPVLTAMAOHTLDGTATVPALRKDRSEETALLTALAQLHVGVC
GVWMSGIFAGTGARRVDLPYPPQHERYWRPVAVSGDVVAGLRSNEHPLLGATLILA
AESGEVLIIRSLQTHPLMPLDHAIGENVFPFGTGLLELAHMAADQACDRVGLTLM
TELVLAEANGTQVQLVGPADAGTRTISIHARPDPDPDPQWQHAGGTLTSGERVVD
FGSTVWPVADAVDLLEGFDYDTGYPAFQGLRSVMQCDGAYVEVALPQVADAGA
FGLHPALLDAVLQSHRIAGVSGADDPQLPFAWRGVSLHAGGASVLRVVRVTKGDDMS

```

```

Query Match 56.4%; Score 138.8; DB 1; Length 84985;
Best Local Similarity 72.8%; Pred. No. 3.1e-09;
Matches 179; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 CTGCGCGGCCGACACAGACAAACAGACACACCCCTCTCGCCTCGTTCGCTCCAC 60
Db 15743 CTGACGGCGCTCGCCCGCCCAAGCGCTCCGAGCGCGTACGAGGACCTGTGCGCACCCGG 15684

QY 61 ATCGCCACCGTCTCGGGCCACACACCCCGCCGACACCATCCCGCCGCGCGCTTCGCG 120
Db 15683 GCCGCGCGCGTACTGGGACACCCACGCCGACCGCGTAGGCGCGCGCGCGTTCGCT 15624

QY 121 GACCTCGCGCTCGACTCCCTCACCGCGCTCGAACTACGCAACCGGCTCTCTCCGCGCACCC 180
Db 15623 GACCTCGCGCTCGACTCCCTCAGCGCGCTCGAACTCGCAACCGACTCGCGCGCGCTACC 15564

QY 181 GCACTCGCGCTCCCGACCAACCGCTTCGCTTCGACGACCCCGACCCCAACCGCTTACCCAC 240
Db 15563 GCACTGAACTCGACCGCGCGCTCGCTCTTCGACCAACCCCGCGCGCGCTTCGCGAC 15504

```

```

BASE COUNT      1884 a  4561 c  4005 g  1931 t
ORIGIN
Query Match      61.8%; Score 152; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTGGCGGCGGACAGACACGAAACAGACAGACACACACCTCTCGCCCTGGTCTCGCTCCAC 60
DB      12230 CTGGCGGCGGACAGACACGAAACAGACAGACACACCTCTCGCCCTGGTCTCGCTCCAC 12289

QY      61 ATGCGCACCGTCTGGGCGACACACACACCCCGACACATCCCCCGACCGCGCGTTCGCG 120
DB      12290 ATGCGCACCGTCTGGGCGACACACACACCCCGACACATCCCCCGACCGCGCGTTCGCG 12349

QY      121 GACCTCGGCTTGCATCTCCCTCACCGCGTCA 152
DB      12350 GACCTCGGCTTGCATCTCCCTCACCGCGTCA 12381

RESULT 6
LOCUS      AX250263                      30000 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION Sequence 3 from Patent WO0168867.
ACCESSION  AX250263
VERSION     AX250263.1  GI:15984065
KEYWORDS   Streptomyces cinnamonensis.
SOURCE      Streptomyces cinnamonensis.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE  1 (bases 1 to 30000)
AUTHORS    Leadlay, P.F., Staunton, J. and Ollivnyk, M.C.
TITLE      Polyketides and their synthesis
JOURNAL    Patent: WO 0168867-A 3 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES   Location/Qualifiers
            source
              1..30000
              /organism="Streptomyces cinnamonensis"
              /db_xref="taxon:1900"
              /note="Nucleotides 60001 - 90000 of the monensin
              biosynthetic gene cluster"
BASE COUNT  4041 a  10632 c  11060 g  4267 t
ORIGIN
Query Match      56.7%; Score 139.4; DB 6; Length 30000;
Best Local Similarity 73.1%; Pred. No. 3.5e-09;
Matches 179; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY      1 CTGGCGGCGGACAGACACGAAACAGACACACACCTCTCGCCCTGGTCTCGCTCCAC 60
DB      6404 CTCTCGTCTGTACAGCCGAGGACAGACCGCACCTCTCTCGACTGTGTACGAGCGAC 6463

QY      61 ATGCGCACCGTCTGGGCGACACACACCCCGACACATCCCCCGACCGCGCGTTCGCG 120
DB      6464 GCGGACCGGCTCTCGGCGACCGGCGACCGCGGTACGCGCGCGCTTCGAG 6523

QY      121 GACCTCGGCTTGCATCTCCCTCACCGCGTCA 180
DB      6524 GATCTCGGCTTGCATCTCCCTCACCGGCGTCA 6583

QY      181 GGACTCGGCTCTCCCAACACACCTCTCGACACCCCAACCCCAACCCCTTCACCCAC 240
DB      6584 GGCCTGGCTGCGCGCGGCGGCGGCTTCATCTTCGCGCACCCGACCCCGCTCGCGGAC 6643

QY      241 CACCT 245
DB      6644 GAAC 6648

RESULT 7
SNA278573/c
LOCUS      SNA278573                      84985 bp      DNA      linear      BCT 17-MAY-2001
DEFINITION Streptomyces natalensis pimarin biosynthetic gene cluster.
ACCESSION  AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS   ABC transporter; cholesterol oxidase; cytochrome P450 monooxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehydratase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE  1 (bases 1 to 84985)
AUTHORS    Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N. and Martin, J.F.
TITLE      A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimarin in Streptomyces natalensis
JOURNAL    Chem. Biol. 7 (11), 895-905 (2000)
MEDLINE    20547809
PUBMED     11094342
REFERENCE  2 (bases 1 to 84985)
AUTHORS    Aparicio, J.F.
TITLE      Direct Submission
JOURNAL    Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES   Location/Qualifiers
            source
              1..84985
              /organism="Streptomyces natalensis"
              /db_xref="taxon:68242"
              <1..2086
              /note="ORFX"
              /codon_start=2
              /transl_table=11
              /product="hypothetical protein"
              /protein_id="CAC20917.1"
              /db_xref="GI:12055068"
              /translation="TRPDAPREATEPYGALLAPLMERSADRVETKWTSDWSQARPL
              AHDAVCEVLTLAGQHPVLVLLLEDLQWADTASLDLRLLSPRCGHPGIGIVLTARKFE
              TESDATLRIVSDVLRGRTATRLICGLPRQAVGALVEAHAGSDVPVVEVLRHSE
              GNPFVQLLSLVGDARDHPDADLAIPTGVREALNQRFAALPEVPLVRLC
              AVIGAEDTDLRLTATCDASVAGLESATBAGLGDHPHGHQALFAAEVLPRLLLAA
              ELPRADRRRLHAKVADALYARRRGHVGDEIERVAHSHAKSLPAAEVLPRLLLAA
              EHAEQMAEVEIWLRAVHLVGFPPDDPATVRLQKLIQGVLTATRGYGHPE
              AETALTRGALSTVTHAPEDPSVLWALCAASLVTRGYDASRQLSGLRLDLSMOTREP
              AVLGAAYGEGIVLHVRGMLPQALAEHGVHRAHFAHEGHALARTFQHDPRVSCRSY
              NTFTWLLGHFQIADARRRLLSLTEYESRPSDRSFALYVDGVVAAGSDRTALASS
              DEGVRAGDHGLLYKAMLVGVKGMVLHAGDRDEGLALMHTSLAELGQSRTYLRHPL
              HLGLLGQAQHNGQIEEARTTFRMLAAVEQRRERVYLRHLPALPATPLHLLHGHGADE
              ALADRDPRGQV"
              2190..3565
              /gene="pimK"
              2190..3566
              /gene="pimK"
              /function="putative glucosyl transferase"
              /codon_start=1
              /transl_table=11
              /product="pimK protein"
              /protein_id="CAC20918.1"
              /db_xref="GI:12055069"
              /translation="MESARRPILFVSLPESGLNPLVLAGELSRQGVDELWLFATDEP
              RRDVNR TAEGSPVEFASLGEVDSMSAVTWSDEVYREVTPQSRPKAHRAVVRHTYRP
              GLQAEKFRLLQAVIDEVQPALMVIDICLSGFADVAIAENI PVLVSPVLPNSVLTHT
              HPKSYTPRGFPVPHITGLSRRTMLAQVRVANELFKLRTFAMFLNPRLGKVLAEADNRNRN
              ELGLPKAFPMARI EHADLVLCNSLAELDYDFDIPEKMLRVGAMVPLPEADQDLRSR
              WLDAQSVVYVGLGTITRLTREQVSMVEARRLEHQLVLPVSEQHLLPRESL
              PGNLRVSWPVSQMDVLAPHVKVFTHGGGNGFNMGYFGKLPVLRVLDYDQAV
              RGQDGLSLTLDRPQTIDVNDVDKLTVLCTPSPYEKERRAALMRSAGGRETAAGL
              VLSLPALA"
              complement(3662..9736)
              /gene="pimS4"
              complement(3662..9736)

```



```
THE KITASATO INSTITUTE
OS Streptomyces avermitilis
PN JP 2000245457-A/1
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR SATOSHI OMURA, HARUO IKEDA
PI C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
PC C12Q1/68//C07D493/22
PH Key (C12N1/15, C12R1.1.465), (C12N9/88, C12R1.465), C12N15/00 CC
FT CDS (1), (11916)
FT CDS (11971), (30687)
FEATURES
    source
    Location/Qualifiers
        1..30690
            /organism="Streptomyces avermitilis"
            /db_xref="taxon:33903"
            /protein_id="BAA8474.1"
            /db_xref="GI:5902891"
            /translation="MQRMDGGEPRPAAGEVLGVADENADGGVVPVPCGGQPMQMGRL
            ELLDASDFRESVRACEAFAPYVDMSEVQLRSDSPDAPGLDRVDVQPTTFAVNMISL
            AALWRSQGVFPCAPWEDNPARLAAVAVNGPRSTVVGAREAVADLADLAAQVTR
            MIPVDVPAHSLMVAIBERVVSGLLPITPRPSRIPFHSVTGGRDLTRELDAAYVRN
            MSSTVRPEPAELLQQGPKTFVEMSPHVLTMGLQELAPDLGDTTGTADTVMCTLR
            RGQTLHFUTSLAQLRGHGTSATTVLSARLTALSPTQQQSLLLDLVRANTMAVLND
            DGNRTASDAGSPASFALHGFDSVMGVELNRSLKATGLRPLVTLPIPHITPAAVAAR
            LRTAALGHLDEDTAPVDPSPSGHGTAAADPIAIIIGMACFPVGGVSRPKDLWELAA
            GDAIGPPTDRGWPTQORHAQDPTQGTFFQGGFLHDAHFADGFGISPREALA
            MDPORLLETSWEAFERAGIDPLSVGRSRTGVFAGLSFDYPRMDPTASSEGADVE
            CHILTTGVSLSGRIAYSGLEGAITVDTGCSASLVTLHLAQSLSSGRCSTLALAG
            GVSMTLGMFIERSRQGLSVDRGCKAYSAADAGTGWGEGVGMMLVERLDARVLGH
            RVLAVWVGSGVDMGASNLTPAPNGPQERIRQALANAGLSVADVDVEHGTTGL
            GDPTBAQALLATYQGRAGDRLWGLSKNSIIGHTMAAGVGGVIGMNAALREGVLPT
            LHVDPSPQSDVAGAVRLLEAVPMPGDAAGRLRRAGVSGFISGTTNAHVILSEAPA
            AGCAGGGLVLEGAGLAISVAESVAPVAVPAESVPPVPPVPPVPPVPPVPPVPPVPPV
            LRAQEAALQVAVRPPVSLADVAGLACGRAVLEHRAVVAADRELVQGLGALAAQ
            EPDRVTTHGAPGGDRGGVVPVPCGGQWAGMGVRLASSPVPARNQACEEALAPW
            VDWSDILRRDAGDAWVERADVPQVLFVSVLSALWMSIGIEPDLVLAQSGEIA
            AAHVCGALSALDAKTAVALRAALAAVRGGMASVPLPAQVEPOLIGERWAGRLWVA
            AVNGPSTASVGDAADEVILAYCAGTGVRARRIPDYASHCHVQPLURELLELLGD
            ISPOSGVPPFTVEGTWLDITTLDAAYWYRNLPQVRFSDAVALADDGHRVFEVS
            PHPTLPAIEDTETEDTAEDVTAIGSLRRGNDTRFELTALAHHTTGITGTPTTWHHY
            THHTHPHPTHLDLPTYPFOHQHYLESSQPGAGSGSAGAGSGAGSGRAGTAGTA
            EVESRFDVARQDLVATVLPVPSAGLDTVPFALSAMHRRHQDARINTWTQET
            WKPLTLPTTHQPHQTLIAIPETQTHHPHITNLTLNHLHGTITPLTLNHTHTQTH
            LHHTLHHTRQQAQNHGTGAIITGLSLDALDETTPHHPHTTGLLNLTLTQTHPT
            PPTLVATNTATTHPNDPLTHPTQACTWGLTLLLEHPTHTAGIIDLPTTPTPT
            LOHLTOTLTOPHOTQAIARTTGTHTLRTLTPTTPTHTPTPTPTPTPTPTPTPTPT
            ATHLHLTHQTHQTHLLTSRTGPHTHQTHLLTQLOQKGIHLTITTCDSNPDQLQ
            QLLNTIPPQHPHTTTHAGILDATLNTLPTQLNVLRAKASAHLLHQLTQHTPL
            TAFVLYSAAATFAGPQQAAYANAYLDALAHRRHHLHPATSIAMGTWQGNGLADS
            DKARAYLDRGRFRMSPPELATAVTOAIADTERPVVVIADIDWSKIETSDILVSA
            AREREPVQRPPTPAELHKTUHOTSDORAALELVVRDHVAALPKAIAPDQS
            FRALGDSLTAVEERNLLIKATGLRLVSLVDPHTPAKLAVHLQNLQRJTAABEAPS
            AAATASNSVTEPIAVGMACRPFPGCVTSADDLWDLISSEQDAGGPTDRGWLDLTL
            YDPPDHPGTCTYRNGGLYDAGHDAEFISPREALAMPQQLLLETWETLLEHA
            GINPHTLHGPTGTGTNGQYALRVHNAQSGTDGFTALTGTAGSVISGRISYTFGE
            GPVAVSDTACSSILVALHACRAGECSMALAGVTMSPGAFVFSRQGLAAD
            GHKAFSAADGTCMGEGVGMMLVERLSDAHRNGHRLVAVRGSANVDQNSGLTAP
            NGPSQORVIRQALANAGLSAGDVDAVEAHGTGTLGDPIEAQALATYQDRAAGTGL
            WLGSVKSNGVHTQAAAGVAGVIMVMAALRHGLLRLPTLHVDEPSHYDMSAGAVQLTE
            TYPMPGEGRURRAGQSVFSGVGNNAHVILEEAPADDVPGPPAGEGSDGDEAAG
            SPGVMPMLVSAKSPALRAQALHHLTHDPLGLDLDADVGYTLAHARAFPHRATLIA
            ADRTFLQALQALAADEPHPAVTHSSAPGCTGTGEAAKTAFCISGQAGNPGMAAGL
            YTHPVFAAALNDICTHLDHPLDPLPLLTQNDNEDNAALLOOTRYAQAPAFQ
            VALRLITDGVIHTPHYVAGSLGEITAAHLAGILTLDTATTLTORALTQMTMPPGT
            MTUHTTHPHITLHHLTAHQDLQATLNTLPTLPTLPTLPTLPTLPTLPTLPTLPTLPT
            LPTNHAHSPHTNPLNLQHQTLTTLTHPHTLPTLPTLPTLPTLPTLPTLPTLPTLPTLPT
            VDYATTTQTLHQGVITYIELGPDNTLTLTHNLNPNPTTLTLTHPHHPHTLHLLT
            NLAKTTTTHPHHYTHNDQPHHTLTLPTYPFOHHHYMLESTOPQAGNVSAAQLDP
            TEHPLIGATLELATDGGALLAGRLSRHPMLADHAGVGTGLLSGATFLELALHAGTY
            VGCDRDELTLHAPLVAVPDGVSQVGAADGREGRLSVYVAGSGACGGGAGG
            VWTCHASGLVLEAAGVVDGLAGVWPPRGAADVVDGDRDLRAGACVGLPVPFSGL
            RAVWRDGGDLLEAEVCLPEEAWGDAAGFGLHLLDGVVQPLSLVLLPGGTGFGGAGF
            EGVVPVAVWVGSLHRAVGTGTVRVRVSAVGRGGREAVSVVVDGDEAGVVASVDRLEL
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 477.19 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658

Perfect score: 246

Sequence: 1 ctggcgccgacacacga.....ccacctcaccacacaccc 246

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sta:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_ot:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sta:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	246	100.0	30690	6	E38020	E38020 Avermectin
2	246	100.0	64957	1	AB032367	AB032367 Streptomy
3	244.4	99.3	31422	6	E38021	E38021 Avermectin
C 4	244.4	99.3	64957	1	AB032367	AB032367 Streptomy
5	152	61.8	12381	6	AX006889	AX006889 Sequence
6	139.4	56.7	30000	6	AX250263	AX250263 Sequence
C 7	138.8	56.4	84985	1	SNA278573	AJ278573 Streptomy
8	134	54.5	30000	6	AX250262	AX250262 Sequence
9	133	54.1	113193	1	AF357202	AF357202 Streptomy
10	133	54.1	123580	1	AF263912	AF263912 Streptomy
11	133	54.1	125401	6	AX211739	AX211739 Sequence
12	130.8	53.2	78210	1	AB070949	AB070949 Streptomy
13	129.8	52.8	65140	6	AX211705	AX211705 Sequence
14	128.2	52.1	43280	1	SFU78289	U78289 Streptomyce
C 15	123.6	50.2	13970	1	AE007061	AE007061 Mycobacte
C 16	123.6	50.2	53450	1	MTV018	AL021899 Mycobacte
17	123.4	50.2	2700	6	AX211733	AX211733 Sequence
18	123.4	50.2	2706	1	AF420594	AF420594 Streptomy
19	123.4	50.2	104326	1	AB070940	AB070940 Streptomy
20	121.8	49.5	1681	6	AR173228	AR173228 Sequence
C 21	121.2	49.3	39314	1	SGR300302	AJ300302 Streptomy
C 22	120.4	48.9	82746	1	AF453501	AF453501 Actinosyn
23	120.2	48.9	25150	1	AB070945	AB070945 Streptomy
C 24	118.8	48.3	24225	1	SC2C4	AL512902 Streptomy
25	118.6	48.2	37948	1	AF079138	AF079138 Streptomy
26	118.6	48.2	41097	1	AF016585	AF016585 Streptomy
27	118.2	48.0	11480	1	STWPK3ORF	L09654 Streptomyce
28	118.2	48.0	24568	1	AF220951	AF220951 Streptomy
29	118.2	48.0	50937	6	AR159871	AR159871 Sequence
30	118	48.0	27541	6	AX211706	AX211706 Sequence
31	117	47.6	20394	1	SNA13222	AJ132222 Streptomy
32	117	47.6	20394	6	AX067996	AX067996 Sequence
33	116	47.2	32870	1	AF007101	AF007101 Streptomy
34	113.8	46.3	1784	1	AF41157383	AF411575 Actinomad
35	112.2	45.6	53784	1	AMM223012	AJ223012 Amycolato
36	112.2	45.6	53789	6	A69720	A69720 Sequence 3
37	112.2	45.6	90445	1	AF040570	AF040570 Amycolato
38	111.8	45.4	9769	1	SEERYA	X56107 S. erythrae
39	111.8	45.4	20444	1	SEERYABS	X63569 S. erythrae
40	110.6	45.0	41782	1	AY118081	AY118081 Streptomy
C 41	110.6	45.0	104326	1	AB070940	AB070940 Streptomy
42	109	44.3	17101	1	AB017641	AB017641 Micromono
43	108.8	44.2	20235	1	SERERYAB	M63677 S. erythrae
44	108.8	44.2	20235	6	AR049368	AR049368 Sequence
45	108.8	44.2	20235	6	AR095529	AR095529 Sequence

ALIGNMENTS

RESULT 1

E38020
LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

THIS PAGE BLANK (USPTO)

Db 253 CGCAGCCCGTGCCAAATCATCGCCGA 277

RESULT 14
BE216343
LOCUS
DEFINITION
HVC_ECB0010D23f Hordeum vulgare seedling green leaf EST library
HVC DNA0005 (Blumeria challenged) Hordeum vulgare cDNA clone
HVC_ECB0010D23f, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hordeum vulgare.
EST.
Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 901)
Wing, R., Close, T.J., Klein, A., Wise, R., Wei, F., Begum, D.,
Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi
D.W., Fenton, R.D., Oates, R., and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected incompatible (Mla6) seedling
leaf cDNA library
Unpublished (2001)
On Jul 3, 2000 this sequence version replaced gi:8903955.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 393
Seq primer: AATTAACCTCTACTTAAGGG
High quality sequence stop: 502.
Location/Qualifiers
1..901
/organism="Hordeum vulgare"
/cultivar="CI16151 (Mla6)"
/db_xref="taxon:4513"
/clone="HV_ECB0010D23f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVC DNA0005 (Blumeria challenged)"
/tissue_type="seedling green leaf"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
C.I. 16151 (Mla6) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were challenged with isolate 5874 (AvrMla6
) of Blumeria graminis f. sp. hordei, and leaves were
harvested 20 and 24 hr post-inoculation and snap frozen;
uninoculated leaves were harvested 20 hr post-inoculation
(Wei, Wise). In the TJ Close lab at the University of
California, Riverside, total RNA was prepared from each
sample pool, equal quantities of all three RNA pools were
combined, poly(A) RNA was purified from the mixture, one
pfu were in vivo excised to give pBluescript SK(-) cDNA
phagemids (Choi, Close). Phagemids were plated and picked
at the Clemson University Genomics Institute (CUGI) (Begum
, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Klein, A., Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.

BASE COUNT 189 a 244 c 313 g 155 t
ORIGIN
Query Match 16.8%; Score 43.4; DB 10; Length 901;
Best Local Similarity 50.7%; Pred. No. 6.9;
Matches 104; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 20 CGCAGCAGCAGTCGCTCTCTGACCTGTGTGGCGCCACACCATGCGGTGTGAACG 79
DB 132 CGAGAGGGGAGGATGATGACGCAATGTTGAAGCGCGGATGACGCGGAGGGCGG 191
QY 80 ACGACGGAAACGAGCGCACCGCTGGATGCGCGCCCATCGCGGAGTTTCGCCACCTCG 139
DB 192 AGGGAGCAGCGGCGGAGGACGCGGAGCGGAGCTCAACCTGCGCGCGCTTCGCT 251
QY 140 GCTTCGACTTCGTCATGCGTGTGCTGACCTGCGCAACCGCTCAGCAGCGCGGCTTCG 199
DB 252 TCCACCCACGAGCAGCAGCTCGTGTGCTACCTTCGAGGAAGTGGCGGACAGC 311
QY 200 GTTGGCCCGTGACGCTCATCTTCGA 224
DB 312 CGACGCCCGTCCCAATCATCGCCGA 336

RESULT 15
CNS006XK/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL066051
AL066051.1 GI:4945019
GSS.
Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 935)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..935
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPCI-98"
/note="end : T7"

BASE COUNT 257 a 170 c 162 g 96 t 250 others
ORIGIN
Query Match 16.7%; Score 43.2; DB 17; Length 935;
Best Local Similarity 32.4%; Pred. No. 7.6;
Matches 83; Conservative 62; Mismatches 110; Indels 1; Gaps 1;

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 1217)

AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes

JOURNAL Unpublished (2002)

AUTHORS 2 (bases 1 to 1217)

REFERENCE 2 (bases 1 to 1217)

AUTHORS Coe,E.C.

TITLE Direct Submission

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of

Missouri, Columbia, MO 65211, USA

FEATURES

Source

1..1217
/organism="Zea mays"
/db_xref="MaizeDB:636162"
/db_xref="taxon:4577"
/clone="PCO106344"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"

/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 186 a 452 c 407 g 172 t

ORIGIN

Query Match 17.1%; Score 44.2; DB 11; Length 1217;
Best Local Similarity 50.2%; Pred. No. 4.7;
Matches 109; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 39 CTTGGAACCTGTCGCGCCACACCATGGCGTGTGAACGACGAGGAAACGAGCGCAC 98
DB 852 CGTGGCGTGTGTCGCGACGCGGTGGCGTCCAAAGGGCGACGCCAACGAGCGGG 911
QY 99 CGGTCGGATGCGCGCCCATCGCGAGTTTCGCCACCTCGGTTGACTCCGTCATGG 158
DB 912 CGTTTCGGCGCCCAACCGCGCGCGGTACAAACGCAACCTGGCGCGCGTCTCCGG 971
QY 159 TGTGGAACCTGCGCAACCGCTCAGCAAGGCCAGCGGCTTCGCGTTCGCCGTGACGCTCAT 218
DB 972 CAACGCGCGCACCCCGCGCGCGCCCGACGCGGACATCGAGTGTACCTCTTCGCGCTCT 1031
QY 219 CTTGGAACACACACCGCGCGCGCGTTCGCGCGCGC 255
DB 1032 CAACGAGAACCAAGACCGCGCGCGACCTCCGAGCGC 1068

RESULT 11

BE494141

LOCUS

WHE1251_F05_K09ZS Secale cereale 467 bp mRNA linear EST 02-AUG-2000
cDNA clone WHE1251_F05_K09, mRNA sequence.

ACCESSION BE494141

VERSION BE494141.1

KEYWORDS EST.

SOURCE

ORGANISM

Secale cereale

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

1 (bases 1 to 467)

; Triticeae; Secale.

REFERENCE 1 (bases 1 to 467)

AUTHORS Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton

R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,

Miller,R., Rauech,C.J., Ross,K., Seaton,C.L. and Tong,J.C.

The structure and function of the expressed portion of the wheat

genomes - Another cDNA library from rye

Unpublished (2000)

COMMENT Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

Email: oanderen@w.usda.gov

Sequence have been trimmed to remove vector sequence and low

quality sequence with phred score less than 20

Seq primer: StrataGene SK primer.

Location/Qualifiers

1..467

/organism="Secale cereale"

/cultivar="Blanco"

/db_xref="taxon:4550"

/clone="WHE1251_F05_K09"

/clone_lib="Secale Cereale anther cDNA library"

/tissue_type="Anther"

/dev_stages="Adult plant before anthesis"

/lab_host="E. coli SOLR"

/note="vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: EcoRI; Site 2: XhoI; Plants were grown in the

greenhouse. Anthers were harvested and pooled from early

meiosis to late meiosis. The tissue, total RNA, and

poly(A) RNA were prepared (Butler, Ross and Gustafson) at

University of Missouri, Columbia. A cDNA library was

made, and the cDNA clones were in vivo excised to give

pBluescript phagemids in the TJ Close lab (Choi, Close,

Fenton) at the University of California, Riverside.

Plasmid DNA preparations and DNA sequencing were performed

in the OD Anderson lab (all other authors)."

BASE COUNT 75 a 167 c 141 g 84 t

ORIGIN

Query Match 16.9%; Score 43.6; DB 10; Length 467;
Best Local Similarity 48.8%; Pred. No. 5.9; Mismatches 124; Indels 0; Gaps 0;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 6 CGCGTGTCTCCCGACGACGACGACGCTGCTCTCTGACCTGTGTGCGCGCCACCAT 65
DB 220 CGCGTGTCTCCCGCGCGCGGACCCGTCCTCATCTCGCGCGCGCGCGCG 279
QY 66 GCGGTGTCTGAACGACGACGAGGACGAGCGCGCGCGCGCGCGCGCGCGCG 125
DB 280 CGGTTCGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339
QY 126 TTTGCG 185
DB 340 GGTCTCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399
QY 186 GCGCACG 245
DB 400 GATGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
QY 246 CG 247
DB 460 GG 461

RESULT 12

BG909521

LOCUS

DEFINITION

TaLr1104H07R TaLr1 Triticum aestivum cDNA clone TaLr1104H07 5',

mRNA sequence.

ACCESSION BG909521

VERSION BG909521.1

KEYWORDS GI:14317197

SOURCE EST.

ORGANISM

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 555)

AUTHORS Cloutier,S., Dong,G. and Walsh,A.

BI553137
LOCUS
DEFINITION
603197725F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5277198 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 765)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11700 row: a column: 07
High quality sequence stop: 757.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5277198"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: Sali-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 149 a 257 c 257 g 102 t
ORIGIN

Query Match
Best Local Similarity 17.2%; Score 44.4; DB 13; Length 765;
Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 28 CAGTCGCTGCTCTGGACCTGTGGCGGCCACACCATGGCGGTGCTGAACGACGGA 87
Db 393 CGGGCGACGTGGTGAGAGGCGCACGAGCTCATGTGATCGTGTGCTCAAGACCA 452
QY 88 AACGACGCGACCGGTGCGATGCGGCCCATCGGCGAGTTTCGCCACCTCGGCTTCGAC 147
Db 453 AGAAGATGATCATCTGGTTTCCAGACATGGTGAAGATGTATCGGAGCTACAGAAG 512
QY 148 TCGGTATGGGTGTGAACTGGCGCAACCGCCTCAGCAAGGCCACCGGCTGCGGTTGCC 207
Db 513 TGGTGACGAGCATCTCCGGCGCACACGCTCATCGCCCGGGTGTTCGGGTGTCAC 572
QY 208 GTGACGCTCATCTTGACACACACCGCGCGCGGTGCGC 249
Db 573 CTGAGGCTAACGAGCTGCACACCATGGAGTTTGGCGCTGTC 614

RESULT 9
LOCUS
DEFINITION
BI600118
603246515F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5289003 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 945)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11730 row: m column: 04
High quality sequence stop: 847.
Location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5289003"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: Sali-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.3 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 190 a 306 c 306 g 143 t
ORIGIN

Query Match
Best Local Similarity 17.2%; Score 44.4; DB 13; Length 945;
Matches 111; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 28 CAGTCGCTGCTCTGGACCTGTGGCGGCCACACCATGGCGGTGCTGAACGACGGA 87
Db 373 CAGGCGCAGCTGGTGCAGAGCGGCACGAGCTCATGTGCTGTGCTCAAGACCA 432
QY 88 AACGACGCGACCGGTGCGATGCGGCCCATCGGCGAGTTTCGCCACCTCGGCTTCGAC 147
Db 433 AGAAGATGATCATCTGGTTTCCAGACATGGTGAAGATGTATCGGAGCTACAGAAG 492
QY 148 TCGGTATGGGTGTGCAACTGCGCAACCGCCTCAGCAAGGCCACCGGCTGCGGTTGCC 207
Db 493 TGGTGACGAGCATCTCCGGCGCACACGCTCATCTCCCGCGGTGTTCGGGTGTCAC 552
QY 208 GTGACGCTCATCTTGACACACACCGCGCGCGGTGCGC 249
Db 553 CTGAGGCTAACGAGCTGCACACCATGGAGTTTGGCGTGGTC 594

RESULT 10
LOCUS
DEFINITION
Zea mays PCO106344 mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AY106411
Zea mays
mRNA
linear
HTC 25-MAY-2002

QY 143 TCGACTCCGTATGCTGCGCAACGGCTCAGCAAGGCCAGCGGCTCGCGT 202
 Db 279 GGTGCGCGCGGAGGATTTC---CTGGCACTACACCGGCTCGAGCTCGATCCGCGT 335
 QY 203 TGCCC 207
 Db 336 GGCTC 340

RESULT 6

LOCUS BG817425 508 bp mRNA linear EST 22-MAY-2001
 DEFINITION EM1_75_G09_b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
 ACCESSION BG817425
 VERSION BG817425
 KEYWORDS BG817425.1 GI:14188405
 SOURCE EST.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 508)
 REID, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
 An EST database from Sorghum: developing embryos
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt WM
 LABORATORY for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 TEL: 706 542 1860
 FAX: 706 583 0210
 EMAIL: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 478
 POLYA=No.

FEATURES

source
 1..508 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Embryo 1 (EM1)"
 /notes="Organ: Embryos germinated for 24 hr; Vector:
 Bluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
 EcoRI; The library was made from poly-A RNA in the cloning
 vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."
 67 a 183 c 174 g 84 t

BASE COUNT 67 a 183 c 174 g 84 t
 ORIGIN
 Query Match 17.7%; Score 45.6; DB 12; Length 508;
 Best Local Similarity 52.0%; Pred. No. 2.2;
 Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
 QY 2 TGACCGCGTGTCCCCACGACGACGAGGAAACGAGCGCACCGCGTCGGATGCCGGCCATCGG 61
 Db 244 TGGCGGTGCGCCCAACAGCGCGCGCTGTTCAGTACCGCGCGCGCGCGCGCGCG 303
 QY 62 CCATGCGGTGTGTGAACGACGACGAGGAAACGAGCGCACCGCGTCGGATGCCGGCCATCGG 121
 Db 304 GCATCGCGGTGCGCAAGAGTACTCTCAGATGGCGCGCGCGCGCGCGCGCG 363
 QY 122 CGAGTTTCCGCCACCTCGCGTTCGACTCCGTCATGGGTGTGAGACTGCGGAACCGCCTCA 181
 Db 364 CGCGCATCGCGCGCGCGCGTGGGTGAGTCCATGCGCGCGTCTCTCGCCACGACGCCA 423
 QY 182 GCAAGGCCACGGGCCT 197
 Db 424 GGGCGGCGCGCGCGCT 439

RESULT 7

LOCUS AW576621 524 bp mRNA linear EST 15-MAR-2000
 DEFINITION UI-HF-BROP-aka-b-10-0-UI.g1 NIH_MGC_52 Homo sapiens cDNA clone
 IMAGE:3076194 3', mRNA sequence.
 ACCESSION AW576621
 VERSION AW576621.1 GI:7248160
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 524)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. Tissue Procurement: Louis M.
 Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 The following repetitive elements were found in this cDNA
 sequence:
 334-381, >GC rich#Low complexity
 Seq primer: M13 Forward
 POLYA=No.

FEATURES

source
 1..524 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3076194"
 /clone_lib="NIH MGC 52"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /note="Vector: p7T3-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (7.4-9.5kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Donald, Ph.D. and M. Bento Soares, Ph.D."
 72 a 172 c 194 g 86 t

BASE COUNT 72 a 172 c 194 g 86 t
 ORIGIN

Query Match 17.2%; Score 44.4; DB 10; Length 524;
 Best Local Similarity 50.5%; Pred. No. 4;
 Matches 108; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
 QY 32 CGCTGCTCTTGACCTGGTGGCGGCCACACCATGGCGGTGCTGAACGACGGAACG 91
 Db 380 CGCGCGCGCGGAGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 321
 QY 92 AGCGCACCGGTGGATGCGGCCCATCGGCGAGTTTCCCGCACCTCGCTTCGACTCG 151
 Db 320 AGCTCCCGGACTTCGCTGCCGACACCGCGCGGATCCGGAAGACGACTGGCGCGTGGCGC 261
 QY 152 TCATGGGTGTGCACTGCGCAACCGCTTCAGCAAGCGCACCGCGCTCGGTTGCCGTGA 211
 Db 260 CTTTCCCGGAGCTGTGTGACCGCGCGGTGGAGATCACCGGCCGCTCGAGCGCGA 201
 QY 212 CGCTCATCTTGCACACACACCGCGCGCGCGCGGT 245
 Db 200 TGATCATCAACGCGCTCAATTCCGCGCGCAAGT 167

RESULT 8

Qy	S	CCGCGCTGTCCCCACGCAGCAGCAGTGGCTGCTCTTGACCTGTTGGCGGCCACACCA	64
Db	647	YVCCCCCYVCSCBYBCNTSCSYYSBMSGSSSSSSCSGSSBSYSCCBCCBGYC	588
Qy	65	TGGCGGTGTGAACGACGACGGAACGACGCGCTCGATCGGTCGCGCCCATCGGCA	124
Db	587	GCVSBVCSSTBTSSYSGSBSSTSGTGCKCBSSSSBBCSSBSGCCCYSSYCTYCVS	528
Qy	125	GTTTCGCCCACTCGGCTTCGACTCCGTCATGGGTGTCGAATGCGCAACCGCTCAGCA	184
Db	527	SCCBYCCBGGCBTBSYBSCSGBTBGSBSGKSGKSGGGBKSGBKGYCYCBTYCC	468
Qy	185	AGGCCACGGGCTCGGTGCGCCGTGACGCTCATCTTCGAACACACCGCGCGCGCG	244
Db	467	SBASBCCSSSSSTSKCCSGSSSTKCSBBSBGGSSSKTSGSSSCBSGSSSTGSSSK	408
Qy	245	TCGCGCGCGC	255
Db	407	KYSYSTSSSS	397

RESULT 5	AZ934718	594 bp	DNA	linear	GSS 24-APR-2000
LOCUS	AZ934718				
DEFINITION	BJ_Ba0002F1or B. japonicum BAC library Bradyrhizobium japonicum genomic, DNA sequence.				
ACCESSION	AZ934718				
VERSION	AZ934718.1	GI:13776778			
KEYWORDS	GSS.				
SOURCE	Bradyrhizobium japonicum.				
ORGANISM	Bradyrhizobium japonicum				
	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
	Bradyrhizobiaceae; Bradyrhizobiium.				
REFERENCE	1 (bases 1 to 594)				
AUTHORS	Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.				
	A marker-dense, sequence-ready map of the Bradyrhizobium japonicum genome				
TITLE	genome				
JOURNAL	Genome Res. 11 (8), 1434-1440 (2001)				
MEDLINE	21376150				
COMMENT	Contact: Wing RA				

```

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Class: BAC ends
High quality sequence start: 22
High quality sequence stop: 549.
Location/Qualifiers
1. .594
/organism="Bradyrhizobium japonicum"
/strain="USD110"
/db_xref="taxon:375"
/clone_lib="B. japonicum BAC library"
/lab_host="E. coli"
/notes="vector: pindigo536; site_1: HindIII"
130 a 132 c 171 g 100 t
others
BASE COUNT

```

ORIGIN		Query Match	17.8%	Score 45.8;	DB 17;	Length 594;
		Best Local Similarity	56.8%;	Pred. No. 2;		
		Matches 105;	Conservative 0;	Mismatches 77;	Indels 3;	Gaps 1
Qy	23	AGCAGCAGTCGCTGCTCTCTGGACCTGTGTGGCGGCCACACCATGGCGGTGCTGACGACG	82			
Db	159	AGGAGAAGACGCTGCGGCTCAACCGACGTCGGTATCGACAGTCTGTCGAGCAGTCTTACG	218			
Qy	83	ACGGAAACGAGCGCACCGCGTCGGATCGCGGCCCATCGGCGAGTTTCGCCCACTCGCGT	142			
Db	219	ACATCAACAAAGCGCTGTCTCGCATCAGGCGCGCCTGATCGCTGTCCCGACGCCACG	278			

RESULT 4
 NS00522P/c
 OCU5
 DEFINITION
 CNS0052P. 844 bp DNA linear
 Drosophila melanogaster genome survey sequence TET3 end of BAC #
 BACR1P16 of RPC1-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL036652
 AL036652.1 GI:4932342
 GSS.
 KEYWORDS
 SOURCE
 Drosophila melanogaster.
 Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 844)
 REFERENCE

REFERENCE	1. (bases 1 to 844)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila bac.hcm.

```

FEATURES
source
Location/Qualifiers
1..844
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR11P16"
/clone_lib="RPC1-98"
/note="end : TET3"
261 a 112 c 92 g 35 t 344 others
RIGIN
Query Match 17.8%; Score 46; DB 17; Length 844;
Best Local Similarity 15.1%; Pred. No. 1.9;
Matches 38; Conservative 122; Mismatches 91; Indels 0; Gaps 0;

```

ScorI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same detailed description of the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .925

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR19D16"

/clone_lib="RPCI-98"

/note="end : TET3"

120 a 61 c 61 g 172 t 511 others

BASE COUNT

ORIGIN

Query Match 18.6%; Score 48; DB 17; Length 925;

Best Local Similarity 16.0%; Pred. No. 0.7;

Matches 41; Conservative 126; Mismatches 85; Indels 4; Gaps 1;

QY 5 CCGCGTGTCCCGACGACGAGCAGTCTGCTCTCGACCTGGTGGCGGCCACCA 64

Db 593 CCSSSSGSS 652

QY 65 TGGCGTGTGAACGACGACGAGCAGCAGCGCGTGGATGCGCGGCCACCGCA 124

Db 653 TSKSSSSGSS 712

QY 125 GTTCCCGCCACTCGCTTCACTCGTCACT----GGTGTGCACTGGCGAACGCTC 180

Db 713 SGKSSTBSGSS 772

QY 181 AGCAAGCCACGCGCTCGGTTGCCCGTGGAGCTCATCTCGACACACACGCGGCC 240

Db 773 TSSSSSTGSS 832

QY 241 GCGGTGCGCGCGGCC 256

Db 833 SSSGKGVTGCGGCC 848

RESULT 2

A2934566

LOCUS

DEFINITION

Genomic DNA sequence.

ACCESSION

A2934566

VERSION

A2934566.1 GI:13776626

SOURCE

Bradyrhizobium japonicum.

ORGANISM

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE

Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea

J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A.

A marker-dense, sequence-ready map of the Bradyrhizobium japonicum

Genome Res. 11 (8), 1434-1440 (2001)

JOURNAL

MEDLINE

COMMENT

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: wing@clemson.edu

Class: BAC ends

High quality sequence stop: 539.

Location/Qualifiers

1. .591

/organism="Bradyrhizobium japonicum"

/strain="USDA110"

FEATURES

source

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: wing@clemson.edu

Class: BAC ends

High quality sequence stop: 539.

Location/Qualifiers

1. .591

/organism="Bradyrhizobium japonicum"

/strain="USDA110"

/db_xref="taxon:375"
/clone_lib="B. japonicum BAC library"
/lab_host="E. coli"
/note="Vector: pIndigo536; Site_1: HindIII"

130 a 196 c 176 g 89 c

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 106; Conservative

QY 23 AGCAGCACTGCTGCTCTCGACCTGGTGGCGGCCACACCATGGCGGTGCTGAACGACG 82

Db 107 AGGTGAAGTCTGCTGCGCTCAACGAGGCGGTATCGACAGTCTCTGTCAGCAGCTTACG 166

QY 83 ACGGAACGAGCGACGCGCTCGGATCGCGCCCATCGCGAGTTTCGCCACCTCGGCT 142

Db 167 ACATCAACAGCGCTCTCTCGCATGAGGCGCGCTGATCGGTCTGCCACACGACG 226

QY 143 TCGACTCTCGTCTGAGTCTGCAACTGCGCAACCGCTCAGCAAGGCCACGCGCTTGC 202

Db 227 GCGTGGCGCGGAGGATTTTCTGCGCAACTACCGGCTCGGAGCTCGATCGCGCT 283

QY 203 TGCCC 207

Db 284 GGTCTC 288

RESULT 3

CNS0091P/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TET3 and of BAC #

BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

AL053013

VERSION

AL053013.1 GI:4934461

SOURCE

Drosophila melanogaster.

ORGANISM

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

REFERENCE

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequences

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoss in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. .925

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR19D16"

/clone_lib="RPCI-98"

/note="end : TET3"

120 a 61 c 61 g 172 t 511 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 106; Conservative

QY 23 AGCAGCACTGCTGCTCTCGACCTGGTGGCGGCCACACCATGGCGGTGCTGAACGACG 82

Db 107 AGGTGAAGTCTGCTGCGCTCAACGAGGCGGTATCGACAGTCTCTGTCAGCAGCTTACG 166

QY 83 ACGGAACGAGCGACGCGCTCGGATCGCGCCCATCGCGAGTTTCGCCACCTCGGCT 142

Db 167 ACATCAACAGCGCTCTCTCGCATGAGGCGCGCTGATCGGTCTGCCACACGACG 226

QY 143 TCGACTCTCGTCTGAGTCTGCAACTGCGCAACCGCTCAGCAAGGCCACGCGCTTGC 202

Db 227 GCGTGGCGCGGAGGATTTTCTGCGCAACTACCGGCTCGGAGCTCGATCGCGCT 283

QY 203 TGCCC 207

Db 284 GGTCTC 288

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 330.73 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353

Perfect score: 258

Sequence: 1 ctcgacgcgtgtcccccac.....ccgcggtgcgcgcgcctt 258

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estcov:*

6: em_estpl:*

7: em_estro:*

8: em_hc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	48	18.6	925	17	CNS0091P
2	47.4	18.4	591	17	A2934566 BJ_Ba000
C 3	46.4	18.0	925	17	CNS0091P
C 4	46	17.8	844	17	CNS0052P
5	45.8	17.8	594	17	A2934718 BJ_Ba000
6	45.6	17.7	508	12	BG817425 EM1_75_G0

C	7	44.4	17.2	524	10	AW576621	AM576621	UI-HF-BR0
	8	44.4	17.2	765	13	BI553137	BI553137	603197725
	9	44.4	17.2	945	13	BI600118	BI600118	603246515
	10	44.2	17.1	1217	11	AY106411	AY106411	Zea mays
	11	43.6	16.9	467	10	BE494141	BE494141	WHE1251_F
	12	43.4	16.8	555	12	BG909521	BG909521	TALF1104H
	13	43.4	16.8	827	12	BF622916	BF622916	HVSMBA001
	14	43.4	16.8	901	10	BE216343	BE216343	HV_CEB001
C	15	43.2	16.7	935	17	CNS006XK	AL066051	Drosophila
	16	43	16.7	691	10	BE214360	BE214360	HV_CEB000
	17	42.8	16.6	637	14	BQ640195	BQ640195	he25d09.y
	18	42.8	16.6	651	13	BG913908	BG913908	602812419
	19	42.8	16.6	678	10	BE408592	BE408592	601301983
C	20	42.8	16.6	728	17	AZ933916	AZ933916	BJ_Ba000
	21	42.8	16.6	734	12	BG707799	BG707799	602671142
	22	42.8	16.6	740	10	BG708101	BG708101	602671552
	23	42.8	16.6	748	10	BE274823	BE274823	601122771
	24	42.8	16.6	872	9	AUI22140	AUI22140	AUI22140
	25	42.8	16.6	880	9	AUI19522	AUI19522	AUI19522
	26	42.8	16.6	883	12	BG761724	BG761724	602717918
	27	42.8	16.6	903	9	AL543324	AL543324	AL543324
	28	42.8	16.6	941	14	BQ879183	BQ879183	AGENCOURT
	29	42.8	16.6	944	13	BI757009	BI757009	603028730
	30	42.8	16.6	1151	14	BM924919	BM924919	AGENCOURT
	31	42.6	16.5	675	9	AUI26183	AUI26183	AUI26183
	32	42.4	16.4	566	14	BQ752993	BQ752993	WHE4121_F
C	33	42.4	16.4	567	9	AL504599	AL504599	AL504599
	34	42.2	16.4	557	14	BQ762235	BQ762235	EBR001_SQ
C	35	42.2	16.4	681	13	BM347502	BM347502	MEST278-D
	36	42.2	16.4	956	12	BG837420	BG837420	Zm10_07b1
	37	42	16.3	926	12	BG321314	BG321314	Zm04_04d1
	38	41.8	16.2	477	12	BG354194	BG354194	947032801
	39	41.8	16.2	857	17	AG061653	AG061653	Pan trogl
	40	41.8	16.2	1101	17	CNS017SY	AL108460	Drosophila
	41	41.6	16.1	390	10	BE588154	BE588154	WHE0657_C
	42	41.6	16.1	625	13	BM440334	BM440334	pgrin.pk0
	43	41.6	16.1	629	13	BM487133	BM487133	pgrm2n.pk0
	44	41.6	16.1	634	13	BM487449	BM487449	pgrm2n.pk0
	45	41.6	16.1	676	9	AJ444452	AJ444452	AJ444452

ALIGNMENTS

RESULT 1	CNS0091P	925 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #				
DEFINITION	BAC19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL053013				
VERSION	AL053013.1	GI:4934461			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 925)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial				

QY 421 CTGGGTCGCGCGGAGGCAACTACGCGCGGCAATGCGCTCTGACGCGCTGGCG 480
 Db |||||
 769 NGNNNGNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 828
 QY 481 TACCGCGCGCGCGCGCGGTCGCGCGGCGGTCGCTGCGGTGGGGGCTGTGG 534
 Db |||||
 829 GGG 882

RESULT 15

BI416517 897 bp mRNA linear EST 15-AUG-2001
 LOCUS hasp001x1d10f Heterobasidion annosum - Scots pine infection stage
 DEFINITION subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
 annosum cDNA clone hasp001x1d10f, mRNA sequence.

ACCESSION BI416517
 VERSION BI416517
 KEYWORDS GI:15187340
 SOURCE Pinus sylvestris/Heterobasidion annosum.
 ORGANISM Pinus sylvestris/Heterobasidion annosum
 Eukaryota; mixed EST libraries.

REFERENCE 1 (bases 1 to 897)
 AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
 TITLE Expressed sequence tags of randomly selected cDNA clones from the
 interaction of the root rot fungus (Heterobasidion annosum) with
 seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)
 COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
 Sweden

Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@kopat.slu.se
 Seq primer: T7 primer.

FEATURES

source
 1. .897
 Location/Qualifiers
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp001x1d10f"
 /clone_lib="Heterobasidion annosum - Scots pine infection
 stage subtraction cDNA library (hasp)"
 /dev_stage="Seedling roots of scots pine were infected for
 6 days with H. annosum"
 /notes="Vector: pT-Adv; Site 1: EcoRI; The subtractive
 hybridization cDNA library was constructed from scots
 pine roots infected for 6-days with mycelia of
 Heterobasidion annosum (FP5)."
 BASE COUNT 12 a 21 c 709 g 47 t 108 others
 ORIGIN

Query Match 17.5%; Score 93.4; DB 13; Length 897;
 Best Local Similarity 45.7%; Pred. No. 1.9e-08;
 Matches 244; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

QY 1 GGGTCTGGTGTGTGACGGGTGGGAGCGGTGTCTGCGCGCGGTGGCGCGGCATCTG 60
 Db |||||
 150 GGGGGGGGNGGGGNGGNNNNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 209
 QY 61 GCTGGTGTGTGTGGGTGCGGATCTGCTGTTGGTGAGCGCGGTGCGGATGCTCCG 120
 Db |||||
 210 GGGGGGGGCGGG 269
 QY 121 GGTGCGGAGGCTGTCGGGGGAGCTGCGCGCGGTGGGGCGGAGGTGCGGATTTGCG 180
 Db |||||
 270 GGGGGGGGNNNNNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGGNGG 329
 QY 181 TGTGATGTGGGAGCGCGGAGGTGTCCTGCGCTGCTGAGGGTGTTCCTGCCGGGTG 240
 Db |||||
 330 GTNGGGGNGGG 389
 QY 241 CGGCTGACGGGTGTCTGTCATGCGGTGTGCTGTGACGATCGGCTCTCTC 300
 Db |||||

Search completed: June 18, 2003, 01:00:31
 Job time : 686.626 secs

Db 390 GGGGGGGGGGTGTGG 449
 QY 301 ACGCCGAGCGCTGGGCAACGCTGTTCGCGCGCAAGGTGGATGCGCTCTTTTGTGAT 360
 Db |||||
 450 GGGGGGGGGGTGG 509
 QY 361 GAGCTGACGCGGGGTATGAGCTGTTCGCGCTTCGTGCTGTTCTCTCTCGCGCGCGGATC 420
 Db |||||
 510 GGGGGGTGG 569
 QY 421 CTGGGTCGCGCGGCGAGGCAACTACGCCCGGCCAATGCGCTCTGAGCGCTGGCG 480
 Db |||||
 570 GTGGGGGNGGCGGG 629
 QY 481 TACCGCGCGCGCGCGCGGTCTGCGCGGGGTGTCGCTGGGTGGGGGCTGTGG 534
 Db |||||
 630 TNGGTGG 683

ORIGIN

```
Query Match      17.7%; Score 94.6; DB 14; Length 993;
Best Local Similarity 44.6%; Pred. No. 1.2e-08;
Matches 238; Conservative 0; Mismatches 296; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGACGGGTGGACGGGTGTGCTGGGTGCGCGCGGTGCGCGCATCTG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 444

QY 61 GCTGTGTGTGTGGGTGGGGATCTGCTGTGGTGTAGCCCGCGGTGTCGGATGCTCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 504

QY 121 GGTGCGGAGGCTGTGCGGCGGAGCTGCGCGTGTGGGGGCGAGGTGCGGATGTTGG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 GNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 564

QY 181 TGTGATGTGGGGAGCGCGGGAGGTGGTCCGCTGCTGAGAGGTGTTCTTCCCGGGTGT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 624

QY 241 CCGCTACCGGTGTGTGTCATGCGGCTGTGTGCTGCGATGCGAGATGCGCTCTCTC 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 684

QY 301 ACCTCCGAGCGGTGGGACGGTGTTCGCGGCGCAAGTGTGATGCGCTCTTTTGTGGAT 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 GNGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 744

QY 361 GAGCTGACCGGGGATGAGAGTGTGCGGCTTTCGCTGCTGCTGCTGCGCGCGGATC 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 745 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 804

QY 421 CTGGGGTGGCGCGGCGAGGCAACTAGCGCGGCGCAATGCCGCTCTGACGCGCTGGCG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 GNGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 864

QY 481 TACCGCGCGCGCGCGGCTGTGCGGGGGTGTGCTGCGTGGGGGGTGTGG 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 865 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 918
```

RESULT 12
BM468944/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BM468944 1089 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6447816 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5587245
5', mRNA sequence.
BM468944
EST.
BM468944.1 GI:18517986
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1089)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12356 row: 9 column: 22
High quality sequence stop: 329.
Location/Qualifiers
1. .1089
/organism="Homo sapiens"

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM911719 1184 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6640314 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470806
5', mRNA sequence.
BM911719
EST.
BM911719.1 GI:19362098
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1184)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP

/db_xref="taxon:9606"
/clone="IMAGE:5587245"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-Sport6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 131 a 692 c 87 g 79 t 100 others
ORIGIN

Query Match 17.6%; Score 93.8; DB 13; Length 1089;
Best Local Similarity 45.3%; Pred. No. 1.6e-08;
Matches 242; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGACGGGTGGACGGGTGTGCTGGGTGCGCGCGGTGCGCGCATCTG 60
Db 905 GGG 846

QY 61 GCTGTGTGTGTGGGTGGGGATCTGCTGTGGTGTAGCCCGCGGTGTCGGATGCTCCG 120
Db 845 GNGNNGNNGNNGGG 786

QY 121 GGTGCGGAGGCTGTGCGGCGGAGCTGCGCGCTTGGGGGCGGAGGTGCGGATGTTGG 180
Db 785 GGG 726

QY 181 TGTGATGTGGGGAGCGCGGAGGTGTCGCGCTGTGCTGAGAGGTGTTCTTCCCGGGTGT 240
Db 725 GGG 666

QY 241 CCGCTGACGGGTGTGCTGATGCGCTGTGCTGAGACGATGCCACGATGCGCTCTCTC 300
Db 665 GGG 606

QY 301 ACCTCCGAGCGCTGGGACCGTGTTCGCGGCGCAAGTGTGATGCGCTCTTTTGTGGAT 360
Db 605 GGG 546

QY 361 GAGCTGACCGGGGTATGAGAGTGTGCGGCTTGTGCTGCTCTCTCGCGCGCGGATC 420
Db 545 GGG 486

QY 421 CTGGGGTGGCGCGGCGAGGCAACTAGCGCGGCGCAATGCCGCTCTGACGCGCTGGCG 480
Db 485 GGG 426

QY 481 TACCGCGCGCGCGCGGCTGTGCGGGGGTGTGCTGCGTGGGGGGTGTGG 534
Db 425 GGG 372

RESULT 13
BM911719
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BM911719 1184 bp mRNA linear EST 12-MAR-2002
AGENCOURT_6640314 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470806
5', mRNA sequence.
BM911719
EST.
BM911719.1 GI:19362098
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1184)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP


```

ACCESSION B0939440
VERSION B0939440.1 GI:22354918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1089)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
JOURNAL Email: cgabs@mail.nih.gov
COMMENT Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC2670 row: g column: 02
High quality sequence stop: 327.
Location/Qualifiers
1..1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6484345"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC library."
BASE COUNT 73 a 59 c 712 g 81 t 164 others
ORIGIN
Query Match 17.9%; Score 95.4; DB 14; Length 1089;
Best Local Similarity 44.9%; Pred. No. 8.2e-09;
Matches 240; Conservative 0; Mismatches 294; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTGACGGGTGGGACGGGTGCTCTGGTTCGGCGGTGGCGGCGCATCTG 60
DB 512 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 571

QY 61 GCTGGTGTGTGCGGGTGGCGGATCTGCTTGTGTGAGCCGGCGTGGTCCGGATGCTCCG 120
DB 572 GNGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 631

QY 121 GGTGCGAGAGGTCTGCGGGCGGAGCTGGCGCGTGGGGCGGAGGTGGCGATGTTGTGCG 180
DB 632 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 691

QY 181 TGTGATGTGGGAGCGCGGAGGTGTTCCGCTGCTGAGAGGTGTTCTCCCGGGTGT 240
DB 692 GGGNNNNGNNGNNGGGGGGGGGGGGGNNNNGNNGNNGNNGNNGNNGNNGNNGG 751

QY 241 CCCTGACGGGTGTGTGTCATCGGCTGTGTGTCGTGATCGATCGATCGCTCTCTC 300
DB 752 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 811

QY 301 ACCGCCAGCGGTGGGACGGTGTTCGCGGCCAAGGTGGATCGCGCTCTTTTGTGGAT 360
DB 812 NCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 871

QY 361 GACCTGACGGGGGTATGAGAGCTGTGGCGTTTGTGCTGTTCTCTCTCGCGCGCGGATC 420
DB 872 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 931

```

```

QY 421 CTGGGGTCTGGCGCGGCGAGGGCAACTACGCCGCGGCAATGCGCTCTTGGACGCGTGGCG 480
DB 932 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 991

QY 481 TACCGCGCGCGCGCGCGCGGGTCTGCGCGGGGTGTGCTGCGGTGGGGCTGTGG 534
DB 992 GCGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1045

RESULT 9
AL572700 855 bp mRNA linear EST 16-FEB-2001
LOCUS AL572700 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0D1033YC01 3
DEFINITION prime, mRNA sequence.
ACCESSION AL572700
VERSION AL572700.1 GI:12931222
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 855)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1033YC01"
/clone_lib="LTI_NFL006_PL2"
/notes="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 67 a 95 c 629 g 43 t 21 others
ORIGIN
Query Match 17.8%; Score 95.2; DB 9; Length 855;
Best Local Similarity 47.4%; Pred. No. 8.8e-09;
Matches 253; Conservative 9; Mismatches 272; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGACCGGTGGGACGGGTGTGCTGGTGGCGCGGTGGCGCGCATCTG 60
DB 14 GGGGGGGGGTGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 73

QY 61 GCTGTGTGTGTGGGGTGGGGATCTGCTGTGTGTGAGCCGGCGTGGTCCGGATGCTCCG 120
DB 74 GGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 133

QY 121 GGTGCGGAGGGTCTCGGGCGGAGCTGCGCGCTTGTGGGGCGGAGGTGCGGATGTTTGGC 180
DB 134 CGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 193

QY 181 TGTGATGTGGGAGCGCGGAGGTGTTCGGCTGTGCTGTGAGAGGTGTTCTCTCCGGTGT 240
DB 194 GCGGGGGGGGGTGTGSGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 253

QY 241 CGCTGACGGGTGTGTGTCATCGGCTGTGTGCTGGACGATGCGACGATCGCTCTCTC 300
DB 254 GCGGGGGGGCGGGGGGGGGGGGGTGGCGGGGTGGGGGGGGGGGGGGGGGGGGGGGG 313

```

/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 26 a 54 c 746 g 30 t 233 others
ORIGIN

Query Match 17.9%; Score 95.8; DB 14; Length 1089;
Best Local Similarity 46.8%; Pred. No. 6.9e-09;
Matches 250; Conservative 0; Mismatches 284; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGTCGACGGTGGGACGGGTGCTCGGTGGCGCGGTGGCGGCATCTG 60
Db 374 GGGNGGG 433
QY 61 GCTGGTGTGTCGGGGTGGCGGATCTCTGTGTGAGCCGGGTGTCGGATGCTCCG 120
Db 434 GGG 493
QY 121 GGTGCGAGGGTCTCGGGCGAGCTGGCCGCTGTGGGGCGGAGGTGCGGATTTGTCG 180
Db 494 GGG 553
QY 181 TGTGATGTGGGGAGCGGGGAGGTGTCGGCTCTGAGAGGTGTCGCGGGTGT 240
Db 554 GGG 613
QY 241 CCCTGACGGGTGTGTCGATGCGGTGTGTCGTGACATCGACATCGCCTCTCTC 300
Db 614 GGG 673
QY 301 ACGCCGAGCGGTGGGACAGGTGTCGGCGCCAAAGTGGATGATCCGCTCTTTTGTG 360
Db 674 NGGG 733
QY 361 GACCTACCGGGGTATGAGCTGTGGCGTTGCTGCTGTCTCTCGCGCGGGGATC 420
Db 734 GGG 793
QY 421 CTGGGTGCGCCCGGACGGCAACTACGCGCGCCCAATGCGCCTCTGACGCGGTGGCG 480
Db 794 GGG 853
QY 481 TACCGCGGGCGGGCGGGTCTGCGGGGGGTGTCGTCGCGTGGGGGCTGTGG 534
Db 854 GGNNGGG 907

RESULT 7
BQ943816 1136 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8773466 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6372516
DEFINITION 5', mRNA sequence.
ACCESSION BQ943816
VERSION BQ943816.1 GI:22359294
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1136)
NIH-MGC http://img.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2548 row: 9 column: 13
High quality sequence start: 54
High quality sequence stop: 323.
FEATURES
Location/Qualifiers
1..1136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6372516"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 80 a 268 c 669 g 65 t 54 others
ORIGIN

Query Match 17.9%; Score 95.6; DB 14; Length 1136;
Best Local Similarity 48.4%; Pred. No. 7.5e-09;
Matches 259; Conservative 0; Mismatches 275; Indels 1; Gaps 1;
QY 1 GGGTCGGTGTGTCGACGGTGGGACGGGTGCTGGGTGGCGCGGTGGCGGCATCTG 60
Db 488 GGG 547
QY 61 GCTGTTGTGTGGGTGTCGGGATCTGCTGTTGTGTCGACCGCGGTGTCGGATGCTCCG 120
Db 548 GGG 607
QY 121 GGTGCGAGGTCTCGGGCGGAGCTGGCCGCTGTGGGGCGGAGGTGCGGATTTGTCG 180
Db 608 GNGGG 657
QY 181 TGTGATGTGGGGAGCGCGCGGAGTGTGTCGGCTGTCTGAGAGGTGTTCTCTCGCGGTGT 240
Db 668 GGG 727
QY 241 CCGCTGACGGGTGTGTCGATGCGGCTGGTGTGTCGACGATGCGACGATCGCTCTCTC 300
Db 728 GGGGGCGGG 787
QY 301 ACGCCCGAGCG-GCTGGGGACGGTGTTCGGGGCAAGGTGAGTGCCTCTTTTGTGGA 359
Db 788 CCGGCGGG 847
QY 360 TGAGCTGACGGGGGTATGAGCTGTGCGCGTTCGCTGTGTTCTCTCGCGCGCGGGAT 419
Db 848 GCGGGCGGG 907
QY 420 CTTGGGGTTCGGCGGGCAGGGCAACTACGCGCGCGCCCAATGCGCTCTGAGACGCGTGGC 479
Db 908 GCGCGNNNGGG 967
QY 480 GTACCGGG 534
Db 968 GGGCGCGGG 1022

RESULT 8
BQ939440 1089 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8929085 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6484345
DEFINITION 5', mRNA sequence.

Qy	419	TCTTGGGGTCGCGCCGGCAGGGCAACTACGCCCGCGCCAATGC CGCTCTGGACGCGCTGG	478
Db	600	CCGGGGGGGGGGCGCGCGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGCGNN	659
Qy	479	CGTACC GGCGGGCGGGCTCTGCGGGGGGTGTCGTGGCTGGGGGGCTGTG	533
Db	660	NGGCNCGGGGGGGGGCGGGCGGGCGGGCGGGCGGGCGGGGGCGGGGGCGGGCG	714
RESULT 4			
BQ940345			
LOCUS			
DEFINITION	BQ940345 949 bp mRNA linear EST 21-AUG-2000		
ACCESSION	AGENCOURT 8806832 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6338352		
KEYWORDS	5', mRNA sequence.		
SOURCE	BQ940345		
ORGANISM	BQ940345.1 GI:22355823		
REFERENCE	EST.		
AUTHORS	human.		
TITLE	Homo sapiens		
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
COMMENT	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. NIH-MGC http://mgi.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rgs@nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: Agencourt Bioscience Corporation found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LCM2531 row: h column: 01 High quality sequence stop: 271. Location/Qualifiers 1..949 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6338352" /clone_lib="NIH_MGC_47" /tissue_type="neuroblastoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." Note: This is a NIH_MGC Library." BASE COUNT 55 a 62 c 563 g 86 t 183 others ORIGIN Query Match 18.0%; Score 96.2; DB 14; Length 949; Best Local Similarity 42.8%; Pred. No. 5.7e-09; Matches 227; Conservative 0; Mismatches 306; Indels 0; Gaps 0 QY 1 GGGTCGGTGTTGCTGACGGGTGGGACGGGTGCTGCTGGGTGGCGGGGTGGCGCGCATCTG 60 DB 373 GGGNNNNNNNGGG 432 QY 61 GCTGGTGTTGTTGGGTGCGGGATCTGCTCTTGGTGAAGCCGGCGTGTGTCGAGTCTCCG 120 DB 433 GNNNNNNNGGG 492 QY 121 GTTGGCAGAGGTCTGGCGCGGAGCTGGCCCGCTTGGGGCGGAGGTGGCGATTGTCG 180 DB 493 GGG 552 QY 181 TGTGATGTGGGGGAGCGCGGGAGGTGGTCCGGCTGCTGAGAGGGGTGTTCTGCGGGGTGT 240		

/db_xref="taxon:169015"
 /clone="hasp001xa01f"
 /clone_lib="Heterobasidion annosum - Scots pine infection
 stage subtraction cDNA library (hasp)"
 /dev_stages="Seedling roots of scots pine were infected for
 6 days with H. annosum"
 /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
 hybridization cDNA library was constructed from scots
 pine roots infected for 6-days with mycelia of
 Heterobasidion annosum (FPS)."
 Heterobasidion annosum 28 t 85 others

BASE COUNT 20 a 28 c 1036 g 28 t

Query Match 18.4%; Score 98; DB 13; Length 1197;

Best Local Similarity 46.4%; Pred. No. 2.6e-09;

Matches 248; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY 1 GGTGTCGGTGTGTCACGGTGGACGGGTGCTCGGTGCGCGGTGCGCGGATCTG 60

DB 487 GGG 546

QY 61 GCTGTGTGTGTCGGGTGCGGATCTCTGTTGGTGAGCCGCGTGTCCGATGCTCCG 120

DB 547 GGGGGTGTGTTGGTGTGTTGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 606

QY 121 GGTGCGGAGGGTCTGCGGCGGAGCTGGCGCGTTCGGGGCGGAGGTGCGGATTTGTCG 180

DB 607 GGG 566

QY 181 TGTGATGTGGGAGCGCGGAGGTGTGTCGGGTCTCTGAGAGGTGTTCTGCGCGGTCT 240

DB 667 GGG 726

QY 241 CGCTGACGGTGTCTGTCATCGGCTGTGTCGTCGACGATCGACGATCGCTCTCTC 300

DB 727 NGGGGGGGGNGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 786

QY 301 AGCCCCGAGCGGCTGGGACAGGTGTTTCGGCGGCAAGGTGATGCCCTCTTTTGTGAT 360

DB 787 GGG 846

QY 361 GAGCTACCGGGGTATGAGCTGTGCGGCTTCGTCGTCGTCGTCGTCGTCGTCGTCG 420

DB 847 GGG 906

QY 421 CTGGGTTCGCGCGGCGGCAACTACGCGCGGCAATGCGCTCTGACGCGCTGCGCG 480

DB 907 GGG 966

QY 481 TACCGCGCGGCGGCGGCTCTGCGCGGGGGTCTGCTGCGGTGGGGGCTGTGG 534

DB 967 NGGNGGG 1020

RESULT 2

BM911600 1478 bp mRNA linear EST 12-MAR-2002

LOCUS AGENCOURT_5614915 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466410

DEFINITION 5', mRNA sequence.

ACCESSION BM911600

VERSION BM911600.1 GI:19361979

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Rubin Laboratory
 CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1966 row: m column: 03
 High quality sequence stop: 272.

FEATURES

Location/Qualifiers
 1..1478
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5466410"
 /clone_lib="NIH MGC 41"
 /tissue_type="amelanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dr priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT. (Life Technologies). Note: this is a
 NIH MGC Library."

BASE COUNT 135 a 253 c 815 g 76 t 199 others

Query Match 18.2%; Score 97.4; DB 14; Length 1478;

Best Local Similarity 42.6%; Pred. No. 3.4e-09;

Matches 227; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTCGACGGGTGGGACGGGTGTGTCGGGTGCGCGGTGCGCGGATCTG 60

DB 499 GGG 558

QY 61 GCTGTGTGTGTGCGGATCTGCTGTTGTGTGAGCGGCGTGTGTCGCGATGCTCCG 120

DB 559 GGG 618

QY 121 GGTGCGGAGGGTCTCGGCGCGGAGCTGCCCGTTTGGGGCGGAGTTCGGATTTGTCG 180

DB 619 GGG 678

QY 181 TGTGATGTGGGGAGCGCGCGGAGGTGCTCGCGCTGCTGAGGGTGTCTTCTCCCGGCT 240

DB 679 GGGNGGG 738

QY 241 CGCTGACCGGTGTCTGTCATGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300

DB 739 GGG 798

QY 301 AGCCCCGAGCGCTGGGACCGGTGTTCCGCGCAAGTGGATGCCCTCTTTTGTGAT 360

DB 799 GGG 858

QY 361 GAGCTGACCGGGGTATGAGCTGTGCGGTTCGTGCTGCTTCTCTCGCCCGCGGATC 420

DB 859 GGG 918

QY 421 CTGGGGTTCGCGCGGACAGGGCAATACGCGCGCAATGCGCTCTGTGACCGCTGGCG 480

DB 919 GNNNGGNGGG 978

QY 481 TACCGCGCGGCGCGGGTCTCGCGGGGTGTCTGCTGGGTGGGGCTGTG 533

DB 979 NNNNGGG 1031

RESULT 3

AZ194927

LOCUS

DEFINITION

SP 1029 Al F04 SP6E Strongylocentrotus purpuratus, purple sea

urChin, sperm genomic BAC library Strongylocentrotus purpuratus

AZ194927 741 bp DNA linear GSS 30-AUG-2000

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 684.535 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142

Perfect score: 534

Sequence: 1 gggctgggtgtgtgacggg.....cgctgggtgggggtgtgg 534

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	18.4	1197	13	BI416470 hasp001xa
2	97.4	18.2	1478	14	BM911600 AGENCOURT
3	96.6	18.1	741	17	AZ194927 SP 1029 A
4	96.2	18.0	949	14	BQ940345 AGENCOURT
5	96	18.0	843	17	AZ200896 SP 1011 A
6	95.8	17.9	1089	14	BQ930965 AGENCOURT

7	95.6	17.9	1136	14	BQ943816
8	95.4	17.9	1089	14	BQ939440
9	95.2	17.8	855	9	AL572700
10	94.8	17.8	937	14	BQ941425
11	94.6	17.7	993	14	BQ958704
12	93.8	17.6	1089	13	BM468944
13	93.8	17.6	1184	14	BM911719
14	93.4	17.5	888	17	AZ189327
15	93.4	17.5	897	13	BI416517
16	93.4	17.5	1362	13	BM556092
17	93.2	17.5	748	17	AZ187748
18	93	17.4	1243	14	BM913931
19	92.8	17.4	877	17	AQ788449
20	92.8	17.4	931	13	BI416575
21	92.8	17.4	1090	17	AG128386
22	92.4	17.3	863	13	BI416699
23	92.4	17.3	905	13	BI416651
24	92.4	17.3	939	14	BQ890606
25	92.4	17.3	1006	10	BE034822
26	92.2	17.3	715	9	AL570062
27	92.2	17.3	943	13	BM415351
28	92.2	17.3	953	14	BQ948564
29	92.2	17.3	1047	14	BQ894173
30	92	17.2	729	10	AW583629
31	92	17.2	795	17	AZ188474
32	92	17.2	1011	17	AG062256
33	91.8	17.2	915	17	AZ195437
34	91.8	17.2	932	17	CNS06KYG
35	91.6	17.2	782	12	BG809582
36	91.6	17.2	798	13	BI416695
37	91.6	17.2	864	17	AG060559
38	91.6	17.2	918	17	AQ895329
39	91.4	17.1	542	12	BG085788
40	91.4	17.1	610	9	AL514195
41	91.4	17.1	737	17	AZ193539
42	91.4	17.1	817	17	AZ196702
43	91.4	17.1	993	17	AG137966
44	91.4	17.1	1015	13	BI416650
45	91.4	17.1	1044	17	AG126158

ALIGNMENTS

RESULT 1	BI416470	BI416470	1197 bp	mRNA	linear	EST 15-AUG-2001
LOCUS	BI416470	hasp001xa01f	Heterobasidion annosum - Scots pine infection stage			
DEFINITION	BI416470	subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.				
ACCESSION	BI416470					
VERSION	BI416470.1	GI:15187493				
KEYWORDS	EST.					
SOURCE	Pinus sylvestris/Heterobasidion annosum.					
ORGANISM	Pinus sylvestris/Heterobasidion annosum					
REFERENCE	Eukaryota; mixed EST libraries.					
AUTHORS	1 (bases 1 to 1197)					
TITLE	Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.					
JOURNAL	Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)					
COMMENT	Unpublished (2001)					
	Contact: Fred O. Asiegbu					
	Dept. of Forest Mycology & Pathology					
	Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden					
	Tel: +46 18 67 15 98					
	Fax: +46 18 30 92 45					
	Email: Fred.Asiegbu@mykopat.slu.se					
	Seq primer: T7 primer.					
	Location/Qualifiers					
	1..1197					
	/organism="Pinus sylvestris/Heterobasidion annosum"					

FEATURES
source

THIS PAGE BLANK (USPTO)

Query Match	35.0%; Score 90.2; DB 1; Length 90445;
Best Local Similarity	.63.0%; Pred. No. 7.5e-06;
Matches 162; Conservative	0; Mismatches 83; Indels 12; Gaps 1;

QY	1	CTGACGCGGTGTCCTCCCAAGCAGCAGCAGTGCCTGCTCTCTGGACTGCTGGCGCGCCAC	60
Db	27778	CTGCGCGGCTGCGCCCGCGGACCAAGCTACGCTGCTCTGGACTGCTGGCGCGCAG	27837
QY	61	ACCATGCGGTGCTGAACAGCAGCAGAAAGCAGCGCACCGCTGCGATGSCGGGCCATCG	120
Db	27838	GTCCGCGCGGTGCTCGGACACGCCGA-----CGCAGCGCGTCCGCGTCGAC	27885
QY	121	GGCAGTTTCCCCACTCTCGCTTCGACTCCGTATGGGTGTCGAACTGCGCAACCGCCTC	180
Db	27886	ACGCGCTTCAAGACAGCGCGCTTCGACTCGCTGACCGCGTGCAGTCGCAACCGCATG	27945
QY	181	AGCAAGCCACAGGGCTCGGTTGCCCGTAGCCTCATCTTGCACACACACACGCCGCC	240
Db	27946	CGACCGCCACCGGCTGAAGCTGCCCGGACGCTGCTTTCGACTCCCGAACCCCGAC	28005
QY	241	CGCGTGGCGCGCGCCT	257
Db	28006	CGCGTCCCGCGCACCT	28022

RESULT 13	AX250263/c				
LOCUS	AX250263	Sequence 3 from Patent WO0168867.	30000 bp	DNA	linear PAT 05-OCT-2001
DEFINITION	AX250263				
ACCESSION	AX250263				
VERSION	AX250263.1	GI:15984065			
KEYWORDS	Streptomyces cinnamonensis.				
SOURCE	Streptomyces cinnamonensis				
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.				
REFERENCE	1 (bases 1 to 30000)				
AUTHORS	Leadlay, P.F., Staunton, J. and Ollivnyk, M.C.				
TITLE	Polyketides and their synthesis				
JOURNAL	Patent: WO 0168867-A 3 20-SEP-2001;				
	Biotech Technology Limited (GB)				
FEATURES	Location/Qualifiers				
source	1..30000				
	/organism="Streptomyces cinnamonensis"				
	/db_xref="taxon:1900"				
	/note="Nucleotides 60001 - 90000 of the monensis biosynthetic gene cluster"				
BASE COUNT	4041 a 10632 c 11060 g 4267 t.				
ORIGIN					

[illegible]

SGATVVTDAQAARLRRLAPHVELLVTDGDDVPEGASHYDELAISBPAPBAADDLELDE
PAWMEVTSCTTGRKGVSTORNCILWSVASCVVPFGPGLSDDDRLVWLPPLPHSLISHIA
CVLSATVVGASVRIADGSAADDYMLRIENASSTFLAGVPTTTHHLVRAARGGSESAPS
LRIGLAGGARGLRSEFTTGVPLDIADESSTFCGATINWPDGARGGSCGLAV
PDGAVYTCDDGLDVPAGEBEGVWGNVWMLGYHNSPEATAAAMRGMDGTGDLAR
RQDGYTRICGRIKELIIRIGGANIHPGEVAVLRTVDGVAADGAVGPHDTFTGVEFVA
VYIPGTPGDPAAALTEKREQOISAYKVPDRILEVAHIPTKASGKRRGLLTDPEPAQLR
YAATHEGOSRHADSEAAALRARLSGLDERAQCELLEDLVRTQAAADVGLQVPDGRGA
FRDLFTSLAIVELANRLTEHGLMLPASAVPDHETPAALAPVRAELLIGITQAVAP
VVAADPGEPIVGNACRUPFGGAVGDPEDMLRULVRAERDAVSEFPDGRDMDLSDLLDP
REAGTYSVYGGFLHDAGEFAGFGFISPREAVAMDPOQLLLETSEWALEANGVDP
IALKGTGTFVGLGMLGGQSGGAVAPLEGGFTVTGVASSVAGSRVSYVYLGEGPAVTV
DTACSSSLVAMHLAAQALRGECSMALAGGVTMATPGSFVFSRQALAPDGRCKAF
AAADGTGSEGVGVVYLERLVARERGHRIILAVLRGAVNODGASNGLTAPNGLSQO
VYRALLAAGLAPSDVDVVEAHGTGTLGDIQAQALLATYGOERKQPLMLGSLKSN
IGHAQAAGVAGVYKVKQALRHETIPLTLHVDKPTLEVDWSAGALILELTEARWPRNG
RPRAGVSFGVSGTNAHLILBEEAPAEVVAAPPELVPVPLVSARSTSLSQAEERLA
SLLEDGTVLETAGALVSRRAVDERAVVAVAGSREEAVTGLRALNTAGSGTPKGVVW
PFQGTQWAGMRELLAESPVFAERIAECAALAPWIDMSLVLDLREGDGLRVDVLO
PACFAVWGLAAVWESVVRPDAVVGHSQGEIAAACVSGALSLEDAKQVALLRSQAI
AELSGRWASVALGEDDVVSRVLDGVEVAAVNGPSSVVIAGDAHALDTELILSSEG
TVRRVADVASHTRHVEDIROTLAETLAGISAQAPAPVPTVTSSEWRDGLDGG
MYNLRNQRFGAATALLEQHTVFEVSAHPVTVPQLSELTDAIGTLRREDGGL
RRLASMSGLFVRGTDVDMTAVPAGMWYDLPTAFYFHRHYLLEPAEBSAGDPLLGT
VYSTPGSRLTAVAQMSRRAQPAWVGLVFNALVEAAILRLGLAGTPVVGELVVDAP
VVLPRGSEVOLIVGECEORRRPIEVFSREADEPWTTHAGHTLAPAAAVPEPAAA
GDATVTVAGLEADARYGIIHPALLDAAVTVVGDILLPSVWTVGSLISAGTAVTVTP
TATGLRUTDPAQCVLTVESVRGTPEVASQGTIDALFRVDWPEIIPLETAEDFLPYE
ATSAEATISALQACADPAETRLAVVTGDCTEPGAAPALWGLVRSQSHSGHPRIVLADL
DDPALVPAVASEGQVRVNGVAVSPRLTRVTPQDAPRLPDEGTVITGCTGTGA
LTRHLVTLVAGVHRLVLSRGEAPELOBELTALGASVAIAACVADRAEQAALRAVRI
PAEHLPTAVHTAGVLDDGVTELPDRLATVRPKVDAARLDELTEADLAAVFLF
SAAVGNPQGAAYAAAEALDALAROENSLDLPASTANGMYATVSGMTEHLGDAD
LRNQIRI GMSGLPADEGMALLDAAIATGCTLVAAKFDVAALRAATKAGGVPVLRGL
ALPRLRAAATSLTRLAGLAEQOALLDVRRAAEVLSHSGAESVSHGRTFKD
AGFDSUTAVELNRALAAATGLTSLPAMI FDKPKPALADHLRAKIFGSAANRPAGIT
AAAEPIATVAMACPRFGGVHSPEDMLRVADGADAVTEFPADRGMIDTDLRHPEDH
EGTVYRKGALDDAAGPDAFEGISPEALAMDPOQLLLETSELPERAAIDPTTL
AQDGLVAGVFNVDHSMEMHRAAGVGEPLTGGSSVLSGRVAFHFGGPAVTVDT
ACSSSLVALLHMAVQALRGECSMALAGGVWNGTVETTFEFSRQGLAPDGRCKAFAD
GADGTWSEGVGLLVERLSEARQHQHVLAVRGSVNSDGSNGLTAPNGSPSOQR
IRKALGMSGLSDVDVAEAGTGTGLDPIEAEALLATYQGNRETLPLWGLSVKSNL
HTQAAGVAGIKMYNMAMRGVLPRTLHVDRPSSVYDMSAGAVEALLTEARDVNSGHP
RAAGVSSFGITGQNAHVLEVAAPITTPQEPABEPLVPLVSAKTAAEQACQALRA
FLGLDVTGVRPDAAYALATRAQLDHRVAVLASDRAOLCADLAPGSGVGTGTVDG
KLAVLFTGQSQWACHGRELAETFPVFRDAPFASACVANDTHLREPLREVPDDSALL
DQMTYTGALFVATLFRPSFVGWPGGLAGHSIGELAAHVSGLVDLADAGELVA
ARGRLMQALPAGAMVAQATEDEVAFLDGTVCVAANVNGPDSVLSGTEAAVLAAD
ELAGRKTRRLAVSHAPHSPLMEPLMDDFRAVABRLTYRAGSLPVSTLTGELAAALD
SPDYVYGVRNAVRFSDATVLAGOAGASTFLELPGGALAAAMALGTGQPEQSCVATL
RNGAREVDVLTALAEHLVHVGVDKMTTVLDEPATAVGTVLTPTAFQHORFWDVDET
AAVSTVTPPAEPIVDRPVQDVLVRESAAVNLGHRDAGSPDLDRSPKDHGDSLSAV
KURNRUDFTGVELPSTLIPDYPNPAVLAHLRAELLGERPAAPVTHVDSDEPIAI
VGNSTRPCGANSPEELKLVAGEDAVSFFVDRGMDLGDHPDPHAGHSTYRS
GLHDAHQFDAGLFGISPREALAMDPOQLLLETSEWALEAGVPLPSAGDGVYFT
GIWHDYVTRLEVRGQYTMGTGTASSVAGSRVAVYFPGPAPVTVDTACSSSLVA
MHLAQAALRGECSMALAGTVMASPDALFEFSRQGLSADGRCKAYAEAGDGTGMA
EGVGVVYLERLVARERGHRLVLRGSVNDGASNGLTAPNGSPSOQRVIRGALASA
GLAPSDVVEHGTGTALGDPIEQVALLATYQGEQOPLWGLSKSNIGHTQOAAQV
VGYIKIMAMRHIGWMPATLHVDRSTQVDSAGAEVLTEAREWTRTPRPRGVSF
GASGTNRLIIIEGPAAEAVDEEVSVPVLSARSAGLAGAQLAAALVARNESLAG
VAGALVYKTRILNERAVITVARSDEAQDGLQALARGENAPVTVTAGPKGVWVFP
GOGSWMGGRDLDDSSVPAAARIKECAALBQWTDMSLLDVLRGDALLDRVDVVP
ASPMVWGLAAVWTSIGVTPDVLGHSGQEIAAACVSGALSDDAAKVVALSQIAG
ELVGRGMSVADSESDAVALRTPMANRVEAVNSPSSVITAGDQAQDEALEALAG
DGVRVRVAVDASHTRHVEVTAETLAKTLAGIDARVPAIPFYSTVLGTMVQAVVDA
GMYRLNRQOQVRFGPVADLAGLTVFVEI SAHPVQVLPQSEISDDAVVTGSLRDD
GGLRLLAQAELVYRGVADWTAAVPAAGVVDLPTAFDRHFHMLWAEATAEAEAG
DGEFWTAIESQSDVSLAEILLEVPEQORGALSTVPLVLAQRDRRERSTAEKLRVQVT
WQPEREAGVFGGRWLAVPAGTITDALLKELTGQGLDITVRLEIEEASQAELQLRN
VLAEHDLTGLSLLADGGPADAAEITASTLALVQALGDITTSAPLWCUTSGAVNIGI
QDAVTAQAQAVNGLRAVLERLDRWGLVLDLPAAIDARTAQALGLVINGAGGEDOL
AVRRSGVYRRRLVRKVPESATSRMBPRGTVLVTGAEGLGRHASVWLAQSGAERLIV

CDS

TGTDGDELTAELAEFGTTFVEFCADTDRDAIAQLVADSEVTA VVHAADIAQTSSVDDT
LVWDRRAAKVTTA VMLDQLPEDTPLDAVPFVSSITAGTWGGCGQGPAGANAVLDA
GVWRRARGIKATSIANGALDQIGINDAEALAOQLRRGVI PMAPPLAVTAMVOVAG
NEKAVAVADMDWAAFI PAFTSVRSPPLPADLPEAKATLRAAQDDGEGDGTASSLSL
RVPDPAQNRILLKVLGHASTVIGHSGAEGIGRQAPQGVGFPDSLAJURNJNSLHAA
TAGLDPATLLIPDYPTPALVGYLRVELLRADDGDRDDRLRVLAAPVPAPKBAAG
VLDTLGLADLTGTEGTDATTEAAPADAELDALDISGLVQALQGT5"
15550..30759
/function="b-ketoacyl-ACP-synthase, acyltransferase,
dehydratase, ketoreductase, acylcarrier protein"
/note="ORF2; PKS modules 4-6"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CAA11036.1"
/db_xref="GI:2764762"
/db_xref="SPTREMBL:O52789"
/translation="MANOSMRKNMSAPNEQI VDAIRASLKENVRQLQENSALAAAAAE
PVAIVMSACRYAGGIRGDFMRVVSCEADVYTFPEDRGWDEGLYHPDPNPGTTY
VREGAFLOQDAQDFAGFGISPREALAMDPOQLLEWSETLERAGIDPHSHVRGSDI
GVYAVVHODVAPDLSGFEGPMSLERALGTAGVAGSRVAVTLGLEGPVTVDTMCS8
SLVAIHLAQAALRGECSMALAGGTMATPGGFVAPARQALAPDGRCKVAAAADG
SGWAEGVLLERLSVAREKHQVLAIRGSVANDGASNGLTAPNGPAQQRVIRKA
LASAGLTPSDVTEGHGTGTVLGDPLEVOALLATYGGRPDPQPLWGLSVVGH
QAASGAGVIMVQSLRHGTQLPATQHVDAPTQVDWSAGATELLAEGRVNSGHPRR
GGISFGASGTNAMIILEAPDEPPTAPATGTVPLVSAATAAAGLAQAQALAEV
GDVSLADAGTLVSGRAMLSRAVVVAGSHEEAVTGLRALARGESAPGLLGRSGV
GKYVWPGQGTQWACHGRELLDSSEVPAARIAECETALGHWVMSLTDVLRGEADLL
DRVDVQASTAVVWGLAAVWASLGVEPEAVVHSGQEIAAACVSGALSLEDAKQV
LRSQAIAASLAGRGSASVASEEDATARLEPWARVGRVEAAVNGTSSVVIAGDAEALD
EALDALLDQGVYIRRVAVDYASHTRHVEARDALEMLGGIRAQAPVPTVSTVDTG
BALDGLDQGVYRNRQVRQFPAVAELIEQGHVFEVSAHPVQVIMELVDDTE
AVTGTLRREDGGLRLLASAEELPVSGVTVVDSGLVPPSRRVELPTVAPHQHWLQ
MGSATDAVSLGADHPLGAVVLPQSDGLVFTSRLSKSHMPLHAGHAGVLLI
PGTVVDLARGDELGFVLEELVLEAPLVLRGGVGRVAVVSGNETSRAVDV?
SMRDEGDETHATGLLGASLREPSRPFPAAMPAGAEIPDENFTYDLTERGYAS
GAFQGMVWRGDEVFAEVALPDHREDACKFGLHPALLDAALHTNAPNPDSDRSV
LPPAWGLLHAGVASALRVAPGCPDALTFQAADTGGVLTWMSLVLSVRESAAL
ETAAGEERSLFOVDMTEVATAETAHEAVLEAFGEAPLELTSRVLEAVQSWALD
AADEARLVVTRGAVREVTDPAAGAVWGLVRAAOAENPGRILVDTGDGDPGLGVALAS
GEPQLAVNGNAPVRLARATGEVPEAPVSPSGTVLDTGTSGLGLVAKHLVLA
GVRLVLSARGVAAEDLVLTETQGATVSVACDSRDQVQVALLAHRPTVILHRA
GLLDGVTIGALNRERLAGVAPKDAVOHLDLTDGLDADVPVVSARSTAGLAQALRAAP
NYAANAFDGLMAGRAAGLPGVSLAMGLEQADGLTANLSATDQARMSSGVLPMPT
PAEALDIPDGLAAEQALLVPIKLDRLTQCATAGGVBPHLLRGLVRASRVRTAA
ASGGGLVHLKLAGRPAGEQEAVLGIVQAEAAVLENAPELAGQTCRGSGLDPSLT
AVELNRNSAATGVKLPATLVDPDTPVALARHLEELGETVAGAPATVTVADAGE
PIAVGMACRLPGGVMSPPDDUWRVMAEGRDGMSPPFGDRGMDLGLDPSDPERPGTAY
IRQGGFLHEAALPDPFGFISPREALAMDPOQLLLETSEWALEAGIDPTKARGDAY
GVFGSVSTHYDLESINNPABEGFTVTTATAGSVASGRSVYTFGEPVATVDTACSS
SLVAIHLAQAALRGECMTALAGVAVMVGSPIGVIMSRQORQMAEDGRVKA PADGAG
TVLSEGVIVLLESLVARERGHRLVLRGSVANDGASNGLTAPNGSPSOQRVIRSA
LAGAGLQSVDDVVEHGTGTALGEPTEAQALLATYGKSTETPLWGLSKSNIGHTQA
AAGVAAVIMVQSLRHGTQLPATQHVDAPTQVDWSAGATELLAEGRVNSGHPRRAG
VSPFISGTNAHLILEEAPDPTAPADVPVAVPVVTSARSTAGLAQALRAAPLADG
DVSFRTIGALLSTRATLTDRAVVAGSABEARAGLTALARGESASGLVGTGAPNGK
TVWVPGQGTQWACHGRELLAESPVFAERIIECAALQPMIDMSLDVLRGEGELDRV
DVLQAPCAVWGLAAVWASVGVVDPDVLGHSGQEIAAACVSGALSLEDAKQVALLRS
QAIAELSGRGSASVASEEDATARLEPWARVGRVEIAAVNGPSSVVIAGDAEALTEA
EVLGGRVAVDYASHTRHVEDIQTLAETLAGIDAQAPVVPFYSTVAGETVTDAGVVD
GMYVNRNURNGQGPVPAVAELIEQGHVFEVSAHPVQVIMELTDVAVTGLRRDD
GGRRLTLTSMALFVRGVPVDMATWAPPARVLEPTYAFDHQFMLSPPSAPVAGACPL
AGADHPLGAVLPLPQSDGLVFTSRLSVRTHPLADGVPAALVELAVRAGDEACGPV
LADLTVEKLVLPESGGLRVQVIVSGERTVYVQLECAEDWIRNATCHLSATAPAHE
APDFTWPPAGAQDGLWRRGDIIPARVALPEELDAGAPGHIHPLDAAVQVPLADD
EOPAEWSLVLHAAQASALRVLPVFGGLADTGGVLVLTADSVAGLESGAGKTRAG
SLYRVRETVESIDASAVPANTVEVFAEGEELPTRLVLEAVQTMVLTAAADARLVV
TRGARVETVDPAGAAWGLVRAAOAENPGRIFLIDTDCGEIPALTDDBEIPVARGKFF
VPLRTAPSPGNAVFRPDGTVLISGAGALGLVARRLVERHGVRLVLASRRGRDADG
VADVLADAAVSVVACDSRAQVAAALLDHRPTAVVGTAGVTDAGVLETLDRDRLA
TVFAPDAVRHLDLTDRLDPAFVYTVSSVAVPMGAGSGSYAANAFLUGLMAARR
AAGPLGLSLAWLWDQSGMAAGTDEATRNRMSRRGQIOMTQEGMDLFOALSSAE
SLLVPAKLDLRGVRDAAGGVVPHMLRGLVRAGRAQARAAATVDNGLAGLAGLAPA


```

Best Local Similarity 63.4%; Pred. No. 4.9e-06;
Matches 163; Conservative 0; Mismatches 82; Indels 12; Gaps 1;

QY 1 CTGACCGCGCTGCTCCCGACGAGCAGCAGTGCCTCTCTCGACCTGTCGCGCCAC 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30648 CTGACCGCGCTCAGGCA CAGGAAGGCGACGCGCTCTCTGCGCAGATCGCGCCAC 30707
QY 61 ACCATGCGCGTGTGAACACGACGAGCAACGAGCGCACCGCGTCGGATCGCGCCCATCG 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30708 GCGCGCGCGTGTGCGGCGACGCGCTCCGAGACTCGATCCCGAGGACCGCGCC----- 30761
QY 121 GCGAGTTTCGCCACCTCGCGTTCGACTCGGTATCGGTGTGCAACTGCGCAACCGCTTC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30762 -----TTCAAGGACCTCGCTTCGACTCGCTCACCGCGTGGAGATGCGCAACCGCTG 30815
QY 181 AGCAGGCGACCGCGCTGCGGTGCGCGTGGAGCTCATCTCGACACACCGCGCGGC 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30816 AGCGGCGACCGCGCTCGCGTGGCGCGCACCTCTCTTGGACCGACCGCGCGGC 30875
QY 241 GCGGTGCGCGCGCT 257
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30876 GAGTGGCGCGCACCT 30892

RESULT 9
AB070940/c 104326 bp DNA linear BCT 22-JAN-2002
LOCUS Streptomycetes avermitilis oligomycin biosynthetic gene cluster.
DEFINITION AB070940
ACCESSION AB070940
VERSION AB070940.1 GI:15823967
KEYWORDS Streptomycetes avermitilis DNA.
SOURCE Streptomycetes avermitilis
ORGANISM Streptomycetes avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
1
Muram, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
Genome sequence of an industrial microorganism Streptomycetes
avermitilis: deducing the ability of producing secondary
metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
2 (bases 1 to 104326)
Ikeda, H.
Direct Submission
Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel. +81-3-5791-6242,
Fax: +81-3-3444-6157)
Location/Qualifiers
1. .104326
/organism="Streptomycetes avermitilis"
/db_xref="taxon:33903"
complement(1..1134)
/gene="mcma"
complement(1..1134)
/gene="mcma"
/feature="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=1
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
/translation="MTGYSIADANARFQLVSGSTRVAVVVDLPTFRAGLSDAPAA
RLAGVRGVVDSIDMVLFGGLRDLRVVPSLAVDAPALLVLYQLVAEHHGAPQ
GADPALEAVTAVAGTAYLRAAVTAGLDVTVAPRLTCLVGGTTGRENRAKLRTVR
VNTRTANERFGSGRTAPRLVYVTPDGLATLACTIEDIAGATADVRIERLGTTLR
ENGLAGQLLRGHRDPATPRPHPHQDSLDLQARQTERLAKLRARVQPAVDEALLR
LRKAAGDNDVLYPMREALAARATGVECDALRSTWGTGPSAT"
complement(1131..1886)

FEATURES
source
1. .104326
/gene="ppta"
complement(1131..1886)
/gene="ppta"
/codon_start=1
/transl_table=1
/product="phosphopantetheinyl transferase"
/protein_id="BAB69186.1"
/db_xref="GI:15823969"
/translation="MRAHEVPADVPVSSLHPREALVQDAVESRRRBEFATGRHCARRA
LASGVPGLLATPRGAPRWPAAGVCGGAIHCAGYRAAARAGARAGIRAVGIDAEPDRP
LPEGVLEVALPAEQVKEHTARHPGVWGKLLFSAKESVTKAWFPATGAPLGFEEA
LIDLPVSSVTRVRLPTALRAAPBELDLRGLRWLRHGIIVTAVVDPGDRTPHAP
SPYDRTHSAPTGRSADRPGRPTDRATDRPTA"
complement(1910..2875)
/gene="gimX4 homolog"
complement(1910..2875)
/gene="simX4 homolog"
/codon_start=1
/transl_table=1
/product="simX4 homolog"
/protein_id="BAB69187.1"
/db_xref="GI:15823970"
/translation="MSQDRHDGSTAALGGPSHGRLLAASDIHVHRQNRDIVQDLHP
GCABDLIVAGDVAESVADITWALSLLRDFAKVIWAPGNHELTAPGDPBELGERK
YRHLVDLCALGVVTPEDPYFMHGAAGFVAVFLFLYDYSLFPPGAWSGHALRLA
DAGVICTDEFLHPDPYPYSREACRARVATRLDALPDPLPTVLVNHPLVREPT
RVLRHPEPALWCGTATADWHLRVRAAVVHGLHIPTIRCDVDPHHEVSLGYPREM
QRYGRDPWFQILTPAPAPAPVVPVPPGHRTPASPADRRP"
3213..3971
/gene="olmC"
3213..3971
/gene="olmC"
/codon_start=1
/transl_table=1
/product="thioesterase"
/protein_id="BAB69188.1"
/db_xref="GI:15823971"
/translation="MSANRVGDLNVRQHPAPTAATKLVCLPHAGGSASVPLPVSKA
LAPGVDAIOVPGRODRHEKCDNIAELADAVTAVLEPLADRLPLVFGHSMGATLA
FEVALRLERKGIIVLALPASGRAPSRHSDSVHLRGDDLEIAELKTLSGTSEVLGD
BEILRMVLPFARSDYKAAETRYAGGARLTAPIHNFHGFVEDPRVSLDEAQAADHTTG
HPELRSHPGHFLYNDQAPRVIEETARVSSVAS"
4643..7438
/gene="olmRI"
4643..7438
/gene="olmRI"
/codon_start=1
/transl_table=1
/product="luxR-family transcriptional regulatory protein"
/protein_id="BAB69189.1"
/db_xref="GI:15823972"
/translation="MLIGRGQEIIDLRAYSVSRSTGSRVAVLVESSAAGCKSETLED
LVHVESGLIALRATAAEERNTPLSVLSLVNAGAPPDDLVRFRHTMSAAARMS
TEGRDKARRDAAYELMRTCTEVRALAQAPLVAVDDIIHGDADSLRQLLHARHC
RNVPLITFTQSLAGPADRTFGAELLPNLEHRLPRFCQDITADMLAQGRADAG
QGTVERYPVSGGNPLLLRALLEDLPPDDTDAQDVGPAGQAYGRAVLACLDRGG
EKAHEVAGLAVLGAFTHLLSLDLISAALLENLALRAAGTVGTSFOHPAQA
AVLDSLEPSHLRHOBSAALLHAGAPAAALAEHLHEAGDAGAWGVQLQAAEQA
LLADEARRAVSLELAHASADPRGTETIKRLGSVTRVRSAADEHVDALRGFGA
GLLTPTSDTGGAAALLAHGRLEQASVLEQGRPPGDADRTAARRQLAEIRDSVLNR
GAWCATRAWHENPAGPSTGRAGTGLRAWLSGSHAAALAQRTAGTGLLEITHTLD
ATFDVLNVAWVTLLAGVERALHWCDDTFLREAKRRDAPGWEAVFAVRAEGALRGGN
LAEAERDARRCFVLPERHRSALGSAVSylTVQIARKGYEDAAARLSRMADSLPD
TIHMLDYLARGLYMATROYHSALRPDPHEGLRAERWGTDMPLPRTDTAEALLR
LGEREGURLATEQLAWQDGNPRIKGISLRQLGLASEPAARGLLNRAVELSRADD
RLEISRTLFDLADARETROGRQAMVLRARQLAKSCGAQPRDRIIGSPDAAAGSE
LFGPEERNAWGHLSSESRERVAGLASGYTNREISVRLYITMTSTVEQHLTRVRLK
INGREELLAERAGVQV"
7658..10486
/gene="olmRII"
7658..10486
/gene="olmRII"
/codon_start=1

```



```
GTSGSVSGRVAYTIGLEGPAVTVDTACSSILVALHLAVQALRGCECDMALAGVTVVM
ACGPFMEVRSQRLGAADCRKAFADGATGAWAGAGVVLVERLSDFARLGHPLVLA
VCGSAVNQDASGNLTAPSGPOERVIRQALGNARLTADVADVWEAHGTGTLGDPGIE
AQAALLGTGRDRGGRPVWLSLSKNIHQAAQAAGVGVKMWLAWRGWLPRTHLVD
EFSRHVDSAGSGWLLTTERENPGVDPRRAAVSAGVSGTNAHLILLEPDTFAEASA
TTPVRSESAVLDLREAVPVVVSGRSVVVEAGRLAEVVEAGVAGVAVAVT
MAGRSGFVGLVARGAEALRLARLARGLADGDPDAGVTVGVDPTSGTGGGVVLF
PQCGTQVGMGAGLIGSSSEFAASMEACARALSIVHEWDLLEVSGAGLEVDVVQP
VWAVMSLARVLAAGVAVGHSQCEIAAATVAGALSLEDAAAVVALRAGIGR
VLAGKMAAVPLPAGEVEAGLAKVGVVEAVNGPASTVVSGRDRAVAGVACQAE
GVQARLIPVDYASHSRHVEDLKEGLERVLSGIRPRSPRVPCSTVAGSGPQEPVDAG
YFMRNLRNVEFSAVVGLEEGHRRFIEVSAPHLVHAIEQTAESAADRSHVATGTLR
RQDPSHRLTSTAFAWAGCATLTWDPALPGHLLTPTYPENHHYMLDTPPTPTAR
TQOSPDMVRVYTWKALTSSPESHGICRLVAPPTTDCGELLDTTLVLSERGAS
VLEVPICARAEVAELKPSMAGSEBNTTVSLIGLVPSTDAVRTSIALLOAVSD
ICVPAARWALTRAVAVVPGTPDAGLQAGFGRVALELPDIWGGILIDPETEL
TITPETSQPPQTERLPQTPNRRALLEAAVLAGRGEDQVAVRASGIYGRVRSAAA
AGAAWQPSQVLLTIGDMGAIARRLIARRLAAAGERLVTSSRGPEAPGAELABELR
GHGCEVHAACVDAERDALALATYATPPNAVFTAGILDDAVIDTSPESPETVRGAK
VCGEALLHQTADIKGLDAFLVSSVTGWNAGQCAVAAANAALDALAERRAAGLGP
ATSVAMGLWGGGMAAGEESLSRRGLRAMPDAAVDALIGAMGRNDVCVTVVDVDM
ERFAPATNAIRQRGLDPTVPEAREALTAGTTSATPDGAPELARLSMLNETERLKL
VSLVTEAFVLRHNTDAIGAERPFKSGFSLTSLRLNRNLNAGTGLKLPAIVFD
HSPSTALARLLDRITGAGAPAPADEPPLPVAVADDPPVIVGWACRPPGAGTPEA
LWKLVTREEDVICAAPTDRGDLSDSYDPEPGVAGTYVREGGFLHDAABFDBEFGI
SPREAVMDPQRLILLETSYAIERAGIDPHSLHGRSTGVVGLTHQEYASRLHBAPE
EYEGYLLTKGSASVSGRTSYTLGLEGPSLSDITACSSSLVALHNAQAALRGCECDMA
LAGGVWMAAGLGVFESFORGLAADGRCKAFADGADGTAWAGAGVVLVERLSDAAR
LGHVLAUVCAAGNQDASGNLTAPSGSQERVIRQALANARLTADVADVWEAHGTG
TRLGPIEAECALLGTGRDRDAECVWLSLSKNIHQAAQAAGVGVKMWLAWRGWL
VHGSQCEIAAATVAGALSLEDAAAVVALRAGLIGRLAGRGAMAAVPLPAGVEAGLA
KWPQGVAVNGPASTVVSGRDRAVAVGACQAEQVQARLIPVYASHSRHVEDLKG
ELRVLTPRPSRPGVPCSTVAGSQEPGFVDGATFENLRNRVFEFSAVVGGLLEQG
HLRFTSVSAHPVLVHAI EOTAEADRSVHATGTLRQDDSPHRLTSTAFAWAHAGTL
TWDPALPGHLLTPTYPENHHYVAVTSPAGVDAAAGRCFQWEDHPFLRGGLPLA
DSGERVFAGRGLSEHDLTHAVSGVLLPGTAFVFPALHAGNATGCRRLBELSVEA
PLVLPAAGGVRVQMRVSAADESGRRRIHSAPEAAVHSAAGSGVWTRHGBGTL
VDPTEPDPDADAMAWPAGERVPEPAELYERFAGLYEYGEAFAGVAVWRHQPDA
BEVLLPDRATAGCFWGPALLDAAALQPIWAGLLEVPEDAVLLPFWQGVSLVATG
ACALVRRLTKAGDGVLSQAADTSGAVALSICALVNRPLARKLVLDTGDAERSLY
RVWEPRLLPAGPPRSWAVLGPADRLACTPGLGQDPGGPTALLPEYRALRKALAA
APRPERAVLPLVSGLGVRVSARSGPFRFALVDGDDRTALPLALAVLDDGAGOLS
PGBEIGDLACAGVWGLVRSARSGPFRFALVDGDDRTALPLALAVLDDGAGOLS
LRAGTARTPVLRTAGTPEQRGPAPDPAGTVLTGATGTLRLARHLAAEHGVRHLL
LLSRGGRAAGADELAAELAGLEABPCFAACDAADREALARVLAEPADRLPTGVTHA
AGVLDGDLTALPERIGTVMRPKADALNHLTSTPLSVFVAFSGAAGTIGRPGQ
ANYAANTFLDALAQHRRAGHLPVAVSLANGWGGATGMTGHLSGTDLRMRSSGIAPM
THQGLAFDRALASADEPLVPMRLDUALVREAREHGPDAVPGPILGLLPARAAY
RQAAPVDPGAPAGGEBTAGLEAGLEARELRLVRAEVSGVILGSGPDAVEP
GBVFLKDGFSLTALVRENRNLGAATGLRFTALVDFRTPTSQVAVSLAAELAGPRDGG
D7AAAFEGLELAAVAGLAEADDLRRDVLRRRLTELAALTPQCRNPSAPAPAPSDIL
DRDLSDNDLDFATIEEQI"
942..3929
/misc_feature
/gene="tylg"
/function="encodes starter module"
942..2156
/misc_feature
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase KS-Q of
starter module"
2571..3557
/misc_feature
/gene="tylg"
/function="encodes acyltransferase domain of starter
module"
3675..3929
/misc_feature
/gene="tylg"
/function="encodes acyl carrier protein domain of starter
module"
3993..8471
/misc_feature
```

Query Match

35.6%; Score 91.8; DB 1; Length 43280;

```
/misc_feature
/gene="tylg"
/function="encodes module 1"
3993..5264
/misc_feature
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 1"
5631..8617
/misc_feature
/gene="tylg"
/function="encodes acyltransferase domain of module 1"
7410..7949
/misc_feature
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 1"
8220..8471
/misc_feature
/gene="tylg"
/function="encodes acyl carrier protein domain of module
1"
8541..13970
/misc_feature
/gene="tylg"
/function="encodes beta-ketoacyl ACP synthase of module 2"
10260..11246
/misc_feature
/gene="tylg"
/function="encodes acyltransferase domain of module 2"
11319..11876
/misc_feature
/gene="tylg"
/function="encodes dehydratase domain of module 2"
12861..13415
/misc_feature
/gene="tylg"
/function="encodes beta-ketoreductase domain of module 2"
13719..13970
/misc_feature
/gene="tylg"
/function="encodes acyl carrier protein domain of module
2"
14351..19945
/misc_feature
/gene="tylg"
/codon_start=1
/product="tylactone synthase module 3"
/db_xref="GI:2317861"
/transl_table=11
/protein_id="AAB66505.1"
/translation="MTAELVATRKRLGALREAREPIAVVAMSCRYPGGVTPPEDLMR
LIANDRAVSGLPDRDGMWDLDALYDPCGPGTSYAREGGFLSHCAGFAEFPFGISPRE
ALAMDPOQLLETSSWEALRAGVATDRAGSRCTGVAGVMYDDYGARVLYCAGAGPP
EDLEYLVNGSAGSIASCRVSTGLRCPANTVNTACSSLSVLHLAVRALRGECMD
ALAGATVLSPTPTLVDFSRQGLAPDGRCKAFADSADGTSWAEGMGLLQRLSDAR
REGREVLAVIRGSVAVNODGASNLTPNGRAQORQVIEDALRDAGVGPQDVDAVEAHGT
GTELDGPTTEAGALLATYGTARTAEPLWLSKNIHQAAQAAGVGVKMWLAWRGWL
RLPRTLHVDRPTTRVDMKEGVRLLTEPVWPGGEPRRAGVSGFSGASTNAHVYLE
SVPAGEPAACRPEDTGCATVSGRGPAAALRAQAARLYDALTGCTGTGQAGQAGP
GTAEVAGALAHARTAFHRAVILGGRNRAELLAGLEAEHHPGPRVVTGTAPATERR
TAFLPSCGSGRAGSGRLYRRHPVFAKALDEVCALSPHLRHRPRLDMFAPSGPFEA
EPLRTEFTQALFATQALPRLHGHGRLAEALCGHSGVEIAAAHAGVLTLPDPAAR
LVAAFRMLQALPAGGAMALAEATAEAPILLERAGELALAAVNGSVSVVSGDEAA
VLELLEOWRAGEAREARLAVSHAFSPMDGMLTQFDRVARTLTTPAPTPIVSTLITG
TPVTEELTADHWVROAREPVRFLDAMRTLADGIDTFFVELPGDGLVSAMARDCAAD
RPDGTTCAGDGETPDLTLPLLRSPVETGDAHPGCFERALATAYAGVPLRIAP
APDASLASAAELPTYARQRTYHWDAPAAPALPAGLDDAGHPLLSAALDLPGRGT
VMTGALSATILPWAADHSHVGRVLPGLTALDLDLALHAAAPRVGELTFEAPLPEDEGV
RLRVLAEPDASGRELVSAGEDGWTHTATVLDTGTTFAGPAPPAAPPGD
AELLDAAEYERFADAGIGYGPAPFRGLRSARDGDAILADVRLPCLAGAEADRYGHP
ALLDAALHTAAALGAHGLPFTTWNGVTLHARGAHARVRLTPAGPDPAVATVPDPA
GRPVPTASLTLRPVTTSQLTAAEAARAPLYRVWMTGLPDTGTARDHTWAVAGGCDL
LPGETPHIPDLASALADTGTAPFRVVLADLRGVTNTPRELASQALALVQWADAAEAA
EGRLVLTFRRAVDIGDVTDPAAATVWGLVRAAOSEHPCCFALLDTPDSPSRQLLPR
VAGTAEQULARDGTLTAPSLTRATLPAGARLPALDGLTITGTSGLSGLAEARHLVTR
HGARRLLITSRGSQPAQGAELVAELAAALGAHADVAACDVAADRALRLDRVPAGHP
LTVLHTAGVLDGGLVLTAGTQORLAARPRKADAVRNHLHETLQGHALSAFITYSSAAG
VLSNAGQSGYAANAAYLDSFAVWRSSRGLPAVSLGWPDCGWSGAGTGTABLR
SGIAPLSAREGLAALDAALAAAGDDUTAPAHLLPIKVDATVTLRGADTVFAVLRDLA
```



```

/ gene="nysG"
/ complement (867. .2694)
/ genes="nysG"
/ function="presumably involved in efflux of nystatin"
/ notes="putative transporter (ABC family)"
/ codon_start=1
/ transl_table=11
/ product="NysG"
/ protein_id="AAF71763.1"
/ db_xref="GI:8050837"
/ translations="MASPDLEERTAPRRELVLGLRPHRSVALVSMGVGGVIL
NAPGILLGRVTLIDAGVLGVGPAPGIDFAIGRILLVLLALVLAASLMACGR
LVASVWRTIHEURDAREKLRPLRHFDRPDAGLISLRLVLLVLLVLLVLLVLLV
LITFSLTLMVLMVIPSFLAVMLLSVPVSAIAARIKRAQPHYAAQMSANGTL
NAHVEEVCCTHALIGKDFRRAAEEERFDACNDVAYRAAKAQPASGAMPVMMFVANL
GVYLVAIGAWKINGTTLGDVQAFILVARQFSQPIVEIASVAGRLQSGIASAQRVF
TLLDAPEQDPPLRPTPARAECRVEFTDVSFRYSDPTLLENLSITVFPGSTVAIIG
PTGAKTITGNLRMPRTFEDSGRILLDGDITATMRDRLDRSFGVLQDTWLFGGIIA
ENIYAGACACRADIEEARATCADRFIRLPGQYDTVLDDSGTSGTVAAGEKOLLTVAR
AFLRAPAVLVDIATSSVDTTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIIVM
DACRIVEQGHDDQLCAQGLYARLHAARTHTPTAGAAAG"
/ complement (2662. .4416)
/ gene="nysH"
/ complement (2662. .4416)
/ genes="nysH"
/ function="presumably involved in efflux of nystatin"
/ notes="putative transporter (ABC family)"
/ codon_start=1
/ transl_table=11
/ product="NysH"
/ protein_id="AAF71764.1"
/ db_xref="GI:8050838"
/ translations="MLRLRLRAQLRPYANATLALVALQVLIQTLTLLPTLGAALIDQ
GVVRGGRITELGVVMGVVALVQIAALGAALAAARTATAMGRDLSALFRILDFP
AREIRGFPSSLTRSVNDVQVQNLAQTFGVLVCAPLMCLGSLVLLAQRDVPILALL
LVALLVAVFCGLLRMGTLIYARMLTDLRLGLRLREAITGVRVRSFVRDDHERA
RFAQTAVLVVRRVGRLLIATMLPVLLLMNGFTVALLTWSHRIDAGRGISLSA
LLSYLSILMSVLMFAFLVSPVRAVCAGRIAEVLDTGSSVAPPAAPOVPRGPAGRI
ELCAARYPGPAEPEVLVDLTVEPGERI AVLGTSGKTKTLLNLVRLADATSGAV
RVGQTDVRELTAALAAAGFVQRPYLFSGTVASNLRGRDADTEELWEALVRAQA
ADPVARMPGLDNEITQGGNVSGGQRQLSLARALLRPEIYFLDDCFSAIDQATDA
ALKUTLVYTAGATITVAQRISAGRDADRIIVLDRGRVVAQGHVLLRTSPTVREI
ALSQLEEEAAHGLAGR"
/ 4714. .5748
/ genes="nysDIII"
/ 4714. .5748
/ gene="nysDIII"
/ function="presumably involved in mycosamine biosynthesis"
/ notes="putative dGDP-mannose-4,6-dehydratase"
/ codon_start=1
/ transl_table=11
/ product="NysDIII"
/ protein_id="AAF71765.1"
/ db_xref="GI:8050839"
/ translations="MSKRALITGITGQDGSYLAELHLSQGYQVWGLIRQANPKRSRV
SLIASLDFIDGDLMDQGLSVADTVQDVVNLCAISFVMSQQAELVTEVNGMV
SLRLFAIRMSVSLSTRVSPRGQIRFYQASSSEMGKAAETPQREITLPHRPSPYG
AKAYGHYITRYRSPFGMYAGMLFNHESPRRQGFVTRKISLAVARIKQGLQDLK
ALGNLADRWGADYGVRAHMLMQDQAGDDYVITGTGMHSVRDAVRIAFHGLGNLW
YGDVLDPLVRPAEVEVLCADSAQAQRLGKDPDVFDTLRRMVMVDSLAQVSRNQ
YGDVLLAANN"
/ 5930. .34363
/ gene="nysi"
/ 5930. .34363
/ function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/ note="polyketide synthase"
/ codon_start=1
/ transl_table=11
/ product="Nysi"
/ protein_id="AAF71766.1"
/ db_xref="GI:8050840"

```

```

/translation="MDNEQKRDYLKJATADLRRTRRRVHKLESAAQEPVAIGMTCR
YEGVRSPEDLWRVTEAGEHGVTPFTDRGDWLEALAAAPTASGGLFADDF
GISPREAVAMDPOQRVWLESAMEAFERAGIDPTSVKGSRTGVFIGMAQDPVGPADG
ABGFOLTGNTSVLSRISYTFGTGPVAVTDTACSSSLVAVHLATQALRAGECTIAL
AGGVIIMSGPCTFIEMGRQGLSADGRCSFGDTAGTGWAGVGILVLERLSDAVRN
GHEIIVAVRGTAVNOGASNLTPAGNPSQQQVIOQALVNARLAAGLIDVDVZAHGT
TLGDVEAQLLATYQNRPADRLLLGSVKSNTLQAARAVSVFSGTNAHTIIEQAP
PRTHAEPTTHVDSOGAVRLITDTDPATGPRRAAVSVFSGTNAHTIIEQAP
EPQEDAAQDDAAGSTPATPVVGVVPLLGRTPDRLAQDAALRAALDTPGTRAP
DILLIAHSLATTRAGFEHRAVILATDHPALTDLGTALADADAPAAAPAMITGTTRAET
RLAVLFTQGAQRGLAGRELAARFPATADALDAFTPHLDRLREVINGTDAALL
DKTAQAQPALFAVEVALVRLIESFGRPHDLACHSVGEI VAAHLAGVLISLADAATLVA
ANGRLMQALPDGCAMIAVOASADVAPLACHEDQVATAVNGSPASVWLSAGEAATVA
LAEQAAADCKRTRLIRVSHAFSPMLPMLDAFRAVEDLTPQLPPLPVSNLTGKPA
TVAQUTLADYVDRHVRVAFADGIDMLRHDTTAFLELPGDGLVSAMAQDCLDADA
DAVTLPALRAGPEPEHTLTALAGLHVSGATLDWTGCFAGTGBARTDLPTYAFORRY
WFKALQGTADLRVSLGGAHPLLSAAVSLADAGTLLTGRLSQTHPWLADTVRG
TLLPGTFALELAARAGDEVGCDRVEILTAAFLLLPEGGQVQVQIWNIDPVSGRT
VNVHARPTDGTTPMATHATVLTITADASRLQSPASS EQGTPGLADPHALDAAGWPL
AGAEPLDGHYDLADGGFGYGVFQGLRAARGDGVVYAEVLP EAGRSDAEAFGL
HPALLDAALHAAPFTGLGERGGGLPFSWEGVSLHAGGATTLRLVRLTPVADDALALTV
ADGTGAPVLSVDSLVRVATQOLDTAAAVARDALPRLDWTVPQDTATDGPVALLGA
DPFGLTTHAGFADAPAYFDLAAALADGVPVTVLVLSLGTGDDADPARSAHRCAAE
ALAAVQTLDDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSQSENPGCFVDL
DPDGAAGAAALVAALVSGEPQLAVRGDVLVRVLRPLTEVAGADGTGCGVDSG
VSFGEGLVLTGTGGLGVALRHVLAIEVGRDILLVRSRGERAVAGELVAELAGV
GARVRVACDVTDRAAVVELVGHGAVSAVHAAGVLDGDMGVALTGERLSAVLRPKVD
AVHLHEATRGILDLDLDAFVFSLSAGVFGSPQANYAANAFLDALMTRRRAAGLPGLS
LAWPGSGGMTGLTLDVDAERLARSVPPLSVAQGLALFDAAVAGTADTCVPVRLD
LPVSRARGEVPPLLSLIRVARRAAVAGSATAGNALQRLRLDEGRDEMVDLVRG
QVALVGHATGVDGADGAFRDLGFDLSITAVELNRLNTVTGLRLPATLFDFTVTEGT
LATVYLDLGLTDAEVAVPAQAAVADDPIVGMACRYPGVSSPEDLWRLVTEGT
DAVSGFPTNRGWDVESLVHPDPDHPGTSTYRSGGFLHEAGREFDGFQMSPREALATD
SQORLLLESSWEATERAGIDPVSLGRSTGVFAGVWYSDYSNMLASPFEPGQSGSS
PSLAGRVAITLGLGEPATVDTACSSSLVMAWQALRSCEGCGLALAGGVTVMSSTP
AVFVDFARQGLSPDGRCKFADADGVGSEGVGLVLSRQDVARNGHETLAVVRG
SANOQDASNGLITAPNGSPQQRVIRQALASGLTAGDVVVEAHGTGTLGPIEQAQ
LLATVYDREPERPILLGSKNSLGHGTAAAGVAGVIKWMLRHVGVVPTLHVDAAPS
SHVDSGECAVELLSQAAWPETRVRRAQVSGVSGITNVHVI VEQAPGAKAIAAGA
ARRTPGAVPILLSGRSALRGQAARLLHGLIQARPDALVDVALSIATRSRPFQRAA
VQAQDQLIASLGAALADRPDPAVEGEAAGRGRTAVLTFQGGSORAAMGRELHEVQ
PEFAAFDACAFAVDFDPLRLREVVEFDEGDEAAALDGTQTPALFAVEALFRL
VESWGRPDPFVAGHSIGIAAAHVAGVLTLEDAICRLAARATLMQALTGGAMIAIQ
TEDEIAAHDLDITVAIAVNGVQSVVISGDEEAETIATFAERGGTKRLRVSHAQHS
PRMGMLDAFRIVAEGLTVRAPRIPLVSDLTGRRADDAEVCTAEVWVRVRAVFAAD
CVRTLDRGATTFLELSDGGLLTAMAEDTLGDHDAELVPMLRAGAELEAATLAR
LQVRGVVDVMAAYLAGTCAKRTDLPYAFAYWYMPQLPAPAAALADPADQDLWAA
VERGDARELADILGLGEQDLTPLDSLPLTSSWRNGQEKHLLDLTRYRVEVTRLSKP
TAPVLDGTWLLVASDATADDDCPLTGLTALVALVQALADTGAQRLWATARG
ARTADVDAATQVLSVLPDLDERDADDPCPLTGLTALVALVQALADTGAQRLWATARG
AVSTNPADPVTVPVQAAWGLRGVLESHPRLWGLVDLPQVFDERAGQRLAGLAVK
DAPGEDQVALRATGVSGRRILVRHTVEALPTAAETATGTITLITGCTGLGAEVARWL
ARAGAKHLVTSRRGPDAPGAALAELEGVPSVSVACDVADRDALAAVLTALPEE
LPLTGVHTVAGHYGJPLDTSTAEFAGLTAAKAGAHLALLDALLDFLDFLPGESI
AGVSGNQSAVGAANAYLDALHRRARGLAATSVAMPWAEACMAADVSETLRR
QGLGLDPAWTELRRVAVRQDVTVTVADVDMORYA PLTSARPSALIAGIPEVRL
AADRTEDQDAGSEVTVTRVALAEPEQLRLITDLVTRTESATVLGHSADAVPEGRAP
RDVGDSUTAVELKRUGAATGLSPSTMVFDYPTPLELAQILRAEILGAVUEVAGVP
ATGADDEPPIAII GMACRFPGCVSPESQELWDLVASGTDIAI SEFFVNRMGQTHGLDPP
BTLRTSTGTGFGSSVYOEYAGDGTGEMVTVGSSPSVLSRLISVYFLEGPATVVD
EDRLTSTGTGFGSSVYOEYAGDGTGEMVTVGSSPSVLSRLISVYFLEGPATVVD
TACSSSLVALHLAQSLRNGESNLAVAGGATIMTTPNPIA PSQRALAKGRCKAFS
DDAQMTLAEAGVVLVRLSDAQRNGHPVLAVRGSALINQDGSANGLITAPNGPSQQR
VIRQALANARLAPGDIIDALEAGHTGLPGLDPEIAQAUFATYGRDORDPESALLLGSVKS
NIGHTQSAAGIASIVKVMALRHSELPTLHADAPSSHDWSAGVRLTLTQARAWPET
GRPRRAAVSFGISGTVMHVLLEQAPVADTPAEERPAVAPVIAAGVTVVTVARSA
ALRQABERLLHAETVGTALPAAGPLDGLSLGARARFEHRAVVPVVPAGTDLAALR
AVATDQSPVAVRQVADVEGTVFVPGQSQWQSGMSQLLDESVAFAERIAECAAL
AEFTDMSLVDLVRGVGNAPSRLERVDVQVQSFAMVSLALWRSGVLDPADVHSGQ
EIAAVVSGALSULSDGARVVALRSQAIGRALAGRGMMMSVALSVDLVLEPRVFEFGRV
SVAAVNGPRS VVVAEPEALDARLTADDIRARRIADVASHQVDEHLEELLEV
LAEIAPRTSEVPFTSTVTDWLDLTARMADAGVFRNLGRVRFADAVDLAAEYRAFV
EVSSHVPLSMVAQBAIDBAGVPAVAAAGTLRRDQCGTDRFLLSAAEFVRGVVDVWAGL

```

BELSLDCLRLAQCSATTEQERAAVVMARVYKWEIDPRILVSPMLDELVAALRRGDLRS
 GDAAVITKHLAQWDRPPEAVESLSLTLAAAGDAGEPVEAEHLFVRHMLRSCPPLEAS
 VPGNPVTPRTRPRLLRVSPASVAGVGLKVLTEGDDHTVAMAEHLRGCRSTTV
 EALEAALLAVVAGRPDRAAFCWDGLLQOAGTSTGTSVAIILSGIRAEIARLQOGLPA
 VQOYARSALSALSLPGWVAGVSPALALVHSTIAGSMGVAGAWLNAQVDPDMFTRH
 GLYTHARGHSHLAVDRPAAALDDFLTCGLAKEMGMVPTFLPWRITSAAQOLALGN
 PQOARALREQIRPGGSARARGVALRLAATGEPRRPALLRSVNLCSGDQVE
 LRLALDRQALDHVGAAPAKARTARHNAVADACGDALFKRLFRDVEDREGDPDG
 LTGEQGSASLTAAERRVTTALAALGHSNREIGRKLFTIKSTVEQHLTRVYRKLGVNR
 ADGLDLACTNLGAGTTPAMAMASSAVAG"
 10933..11271
 /codon_start=1
 /transl_table=1
 /product="P450-like protein"
 /protein_id="BAB69191.1"
 /db_xref="GI:15823974"
 /translation="MEVGGQTIRAGEGLVCLTSTANRDEGAFTDDELDRDARSHL
 SFGFIHQICIGSLAKAEFQIALETLLRLPGLRLAASFEEELFRDGVGVYVKELPDP
 PCPALUPPDRP"
 11458..29898
 /gene="olmA1"
 11458..29898
 /gene="olmA1"
 /codon_start=1
 /transl_table=1
 /product="modular polyketide synthase"
 /protein_id="BAB69192.1"
 /db_xref="GI:15823975"
 /translation="MACRLPFGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
 VDFDNGFGTSPREKAMDPQORLVLELSWEALDAGLPPARLGGSGAGVFGAIAD
 DYATLSRAGAATAQTTLTNGRGLIANRVSYTLGRLGPTFTVDSGSSSLVAVHLA
 AESLRGSCSALAGVGLLAEASTVAVREFGALSPDGYCTFDARANGVYRGEGG
 VVVLKPLADAIAGDDIHICILAGSANNDDGGELTPDRAGQEAFLTAHYEQAGISIP
 AAVGVLLHGTGTDPVAAAGAVAGVAGPAGOPVLVGVKTNIGHLEAGIAG
 LLKAVLAVHREIPATLHASNPAPMAELGIGVNTALTTPHPEGAGPLVGVVSFGM
 GTCNVHVTETPAGYDAAAPATAGIRPAGTTPAPLIVSGRDEQALRAQKRDHLDIT
 HGPLRVVAHTLAAAPARLAVRLLAARDVDSAGLARLADGTPGADLVRGTVSPG
 ALAFVFGGSGRPGMTABLYRTSAEYAAALDEVCASLDPMLRRLPDLVFAEGSDC
 AALLDRTEFTQALFAVEALFYAEHCGLTPRFLHSGVGLAARHAGVGLSLADAC
 TLVAARGRMDRPGATGMVSVQATDELAIPVSESVTLAALNGPASTVACGEDAVL
 AVAAHWAQRKTKRLRVSHAFHS PHMDGMLTFEPRIAGLTFGAPRIPIVSNETGL
 LTHEARSPEYVRRHARAVRFLDGVRLLEDQGVTLLELPGDVLVSANARCLRDAG
 TVCPVLLRDRTEPALLAALAGLYVRGPMWENTATEAGARRVPLTYAFQRRH
 WLPEFSASRPVNGVAGQAGHAGQAGAGHAGHAGHAGHAGHAGHAGHAGHAGHAGH
 DRLTEPLTDAAARLEADCAERTDAGPVSPPAAAHDDPIVIVGMCRLPGVSRPE
 LLFDHPTDALLAARLEADCAERTDAGPVSPPAAAHDDPIVIVGMCRLPGVSRPE
 DLWRFAAGDIALSALPDRGHEPADPFGEGFLADVAEPDAGFGISPRALAMP
 QORLLLETSEALERAGVDALKURGRTVFGVSPSEYGPRLHETSDADGHVLTGTA
 PSVLGRIAYVLEGPALTVDPTACSSSLVALHAAQALRGCECDLALAGVTVMAT
 GFAEPARQGLAPDRCKAFADAGDTGCEGVGMVAVORLSDAVDRGPRVLAVRG
 SAVNSDAGNGLTAPNGSQORVIRQALASAGLSAADVDMVEAHGTGSLGDPTEAQA
 LLATYQDRPDRPLWGVSKNIGHTQAAAGVAGVIKSVLALRNLGLLPKTLHVBDEPT
 PEVDMSAGAVELLTEGHEPDTGPRAGVSAFISGTNAHVILSQAPSEVEPVTAPA
 DGLLVVWVSARSBALRAQARLADHVRSGRLPADVGLSLAVTRAGLEHRAVLVA
 EDGAEFLAHLDALEAGRTFAGVVEGTTAFPSGQSQBAGMGRELISYHSPVFAAFD
 EYCAFPDGLPALREVPFGAGLDRTEWAQALFALEALFELVSSWGRADVLVH
 SIGELAAAVAGVWSLADACRIVAARGRLMOALPEGCAVVAEAEDELDPDLPGGVSTL
 AAVGPRSLVLSGDEBPVTALQAPAEQGRRTKOLTVSHAFHSARMEPLAEFAETLA
 AVEFRPPIPVSVNVTGEMAGDEFTTPAYVRHREAVRPADGVGNAHAGVRYLIEL
 GPDGUTTLIADHDAGDAAVAPALRGDQPETRTTVTAVGLRVATGUTVDWPAFFA
 PYGARTELTPTAFQHRRYWLEFVAPSSAPPANPLRYRTAWASVPDPPRSLSTWAV
 VVPASGVEDIVAGVTAALRDHGARPVVVEVDPPAGREADVADTLRAALGEDAGGVVS
 LLGTAGSLSGVAAATVLAQVQWATDLGGARLWCLTRGAVSVSPDPLAAPAQOLWGM
 GRVAALHEPRWGLVDLPVAPDDRAWTRLCALVAGSGCEDQVAVRAAGLFARLLGRA
 CTRADAAPAGPSPWRDGTVLVTGGTGALHAQLWLAAGAARHVLITSRGPDADGA
 ABTLARLETGTETVVAACDTRDALABELLGLPADRLPTGVVHAAGVLDGVDLSL
 TDFRFAAAGVPKVTGAHMLHTRDLDSAFVLFSFAGTGLAGQANAAAYALYDA
 LAVHRAQLGLPATAVAGWSGAGMAADTEAARQLQARTGLVPLAPDAALAAALGOVMD
 DGETAVTADVWDVDRFAAGFGPGRPSPLLTGVPEVRRIBERETGRAGAPAEITDRLLGL

Query Match 36.8%; Score 95; DB 1; Length 104326;
 Best Local Similarity 64.2%; Pred. No. 1.2e-06;
 Matches 165; Conservative 0; Mismatches 80; Indels 12; Gaps 1;

QY 1 CTGACCGCGCTGTCTCCCCACGACGACGAGTCTGCTCTCTGGACCTGCTGGCGCCAC 60
 DB 29395 CTGCGCGCGCTGTCTCCCGCGGACGCGGCTCGTACTGGATCTCTGTCGCGAGCAC 29454
 QY 61 ACCATGGCGGTGCTGACGACGACGAGAAACGAGCGCACCGGTCTGGATGCGCGCCCATCG 120
 DB 29455 ACCGCGGCGCTCTCGGTCTGCGTCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 29502
 QY 121 GCGAGTCTTTCGCGCCACCTCGGCTTCGACTCCGTCATGGGTGTGAACTGCGCAACCGCTC 180
 DB 29503 CAGGCGTTCAAGACGCTCGGCTTCGACTCGCTCACCAGCGGTGCAACTCGGCAACCGGTG 29562
 QY 181 AGCAAGGCGACGCGGCTCGGCTGTCCTCGTACGCTCATCTTTCGACCAACGCGCGGCC 240
 DB 29563 CGGACGCGCACCTCTCCCTGACCGTCCCGCGACCTGGTCTTCGACCAACGCGCGGCC 29622
 QY 241 GCGGTGCGCGCGCGCGCT 257
 DB 29623 GCGCTCACCGCGCATCT 29639

RESULT 6

AF263912 123580 bp DNA linear BCT 24-MAY-2000
 Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
 complete sequence.
 AF263912
 AF263912.1 GI:8050835
 Streptomyces noursei.
 Streptomyces noursei
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 1 (bases 1 to 123580)
 Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strøm,A.R.,
 Valla,S. and Zotchev,S.B.
 Biosynthesis of the polyene antifungal antibiotic nystatin in
 Streptomyces noursei ATCC 11455: analysis of the gene cluster and
 deduction of the biosynthetic pathway
 Chem. Biol. 7 (6), 395-403 (2000)
 20334850
 10873841
 2 (bases 1 to 123580)
 Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strøm,A.R.,
 Valla,S. and Zotchev,S.B.
 Direct Submission
 Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
 N-7489, Norway

FEATURES

source
 1..123580
 /organism="Streptomyces noursei"
 /strain="ATCC 11455"
 /db_xref="ATCC:11455"
 /db_xref="taxon:1971"
 complement (46..783)
 /gene="nysP"
 complement (46..783)
 /gene="nysP"
 /function="presumably post-translationally modifies ACP
 domains on PKS"
 /note="putative 4'-phosphopanthetheine transferase"
 /codon_start=1
 /transl_table=1
 /product="NysP"
 /protein_id="AAF71762.1"
 /db_xref="GI:8050836"
 gene
 CDS
 gene

QY 121 GCGAGTTTCGCCACCTCGCTTCGACTCGGTTCATGGGTGTCGAACCTGCGCAACCGCCTC 180
Db 1378 GCGAGTTTCGCCACCTCGCTTCGACTCGGTTCATGGGTGTCGAACCTGCGCAACCGCCTC 1437
QY 181 AGCAAGGCCACGCGCTCGCGTTGCCGTGACGCTCATCTTCACACACACCGCGCGCC 240
Db 1438 AGCAAGGCCACGCGCTCGCGTTGCCGTGACGCTCATCTTCACACACACCGCGCGCC 1497
QY 241 GCGGTGCGCGCGCGCTT 258
Db 1498 GCGGTGCGCGCGCTT 1515
RESULT 5
AB070940
LOCUS 104326 bp DNA linear BCT 22-JAN-2002
DEFINITION Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION AB070940
VERSION
KEYWORDS
SOURCE
ORGANISM Streptomyces avermitilis DNA.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C.,
Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.,
Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.
TITLE Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403
REFERENCE 2 (bases 1 to 104326)
AUTHORS Ikeda, H.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel. +81-3-5791-6242,
Fax. +81-3-3444-6197)
FEATURES
source Location/Qualifiers
1 .104326
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
complement(1..1134)
/gene="mcmA"
complement(1..1134)
/note="methylmalonyl-CoA mutase"
/codon_start=1
/transl_table=11
/protein_id="BAB69185.1"
/db_xref="GI:15823968"
translation="MTGYSTADANARFQLVSGSTRVAVVVDLPTRAGLSDAPAA
RLGVRGVAVDSIDMVFGLRLDRVPSLAVDPAAPALLVLYQLVAEBHGAPQ
RLAGAVRGVFEFLARGVFPPLRSLRLAMVAYCOAELPRWQCLSVCGHHIAGT
GADPALEAVTAGGIEVYLAATVAGLDVTVAPRLTCLVGGTTGRENRAKLATRR
VWTRANERFGSGTRAPRLVYTRPDGLATLACTEDIEAGATALVDRIERLGLTUVL
EHLGAGQLLGRPHGRPATFRPHPHQSDLSLQAROTERLAKLRAWRVQPAVDALLR
LRKAAQGDNDLVLPNREALAARATVGEVCDALRSTWGTGPSAT"
complement(1131..1886)
/gene="bpta"
complement(1131..1886)
/gene="ppta"
/codon_start=1
/transl_table=11
/product="phosphopanthetheinyl transferase"
/protein_id="BAB69186.1"
/db_xref="GI:15823969"
translation="WRAHEVFADVPVSSLHPREALVQDAVESRRREFATGRHCAARRA
LASLGVFPGLIATPRGAPRWAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPRP
LPEGVLEAVAPAEQVQKEHTARHPGVVWVKLLFSAKESVYKAWFPATGRPLGPEEA

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/product="luxR-family transcriptional regulatory protein"

/protein_id="BAB69190.1"

/db_xref="GI:15823973"

/translation="MLASGQEGGVVWVWAGVGVGRTALLDAFLRKATSGARVCAA

AGSATEAEFGVWQLFPPTPLGAGFRLARIGSDAGPLGDALDFDLCOELRR
RPTVIADVDAQADAASIRFLRYLVRRLTAPVMVLTPTGPDALPSPFOALRHLP
QCRILGLTSDRGVTVQVSAQSDRVPRLAADPHALISGNGNPLLVALLHDLRAHR
LSPVSPVTVPSVSSAASAAASAPAAAPAAAFKAAALAWAYRDEPVLV
DVA CGFVLGESATPARVACLVDRGADAVQDMTALDTAGLLAGTFFSALVKGALL
HVDVETRAELERRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMOEGR

LIDDPVSTFRVRLPTALRAAEPELRLHGRWLVHGGIIVTAVVVDGPGDRDTPHAP
SPYDRKTSAPTGRSADRPGRPTDRATDRPTA"

/gene="simX4 homolog"

/complement(1910..2875)

/gene="simX4 homolog"

/complement(1910..2875)

/gene="simX4 homolog"

/codon_start=1

/transl_table=11

/product="simX4 homolog"

/protein_id="BAB69187.1"

/db_xref="GI:15823970"

/translation="MSQDRHDGSTAAALGSPSHGGRLLAVSDIHHVRHQRNDRIVQDLHP
GCAEDWLIAGDVASVADIWALSRLDRFAKIWAQGNHELWTPAGDPDELGERK
YRHLVLCRALGVVTPEDYPVWGHGAGFVAVPLFLDYDFLSPFGAWSKSHALRLA
DDAGVICTDEFLHPDPYPSREAWCRARVATRLARLDLDDLPVLVNVHFLVREPT
RVLRHPEALWCGTTATADMHLRVEAAAVHGHCHIPTIRICDDVPHHEVSLGYPREW
QRYGRPDWPVQILTPAPAPAPVFPVPAPPPGHRRTASPADRRP"

3213..3971

/gene="olmC"

3213..3971

/gene="olmC"

/codon_start=1

/transl_table=11

/product="thioesterase"

/protein_id="BAB69188.1"

/db_xref="GI:15823971"

/translation="MSANRVQDNLVWROFHPAPTAAKLVCLPHAGGSASYFLPVSKA
LAGVDVLAIOYPGQDRHEKCIDNIAELADAVTAVLEPLADRPLVLFHGHSMGATLA
FEVALRERKIGVPLALPASRRAPSRHRDESIVLRGDEIIAEIKTLCSTRTSEVLGD
BEILRMVLPARSIDYKAAETRYAGGARLTAPIHAHFGVEDPRVSLDEAQAADHTTG
HFELRHPGHHFLYNDQAPRVIEETARVSSVAS"

4643..7438

/gene="olmRI"

4643..7438

/gene="olmRI"

/codon_start=1

/transl_table=11

/product="luxR-family transcriptional regulatory protein"

/protein_id="BAB69189.1"

/db_xref="GI:15823972"

/translation="MLIGRGQEIIDLRAYSVSRTSGRAVLSVESAGCGKSEILED
LVHVESLIALRAIAEERNTPLSVLSVNGAPFPDDLVRFFHTMSDAAARMIS
TEGRKARRDAAYELMRTCTFEVRLAQAPLVVAVDDIHGDDADSLRLHLARHC
RNVPVLLTFTOSLAGGPADTFGAELLPQNLHRLPRFCQDILADMAEQGRADAG
QGTVERYFASVSGNPLILRALLEDARLEPDDTVDQDVGPRAGQAYGRAVLACLDGG
EKAHEVAAGLAVLGCAFTHELLSRLDISAALTENRALRAAGIVGTSFQHPAKA
AVLDSPESHRLRHQSAALLHASGAPAAALHHLHAGQADGAWAVGUGQAAEQA
LLADEARRAYSILEHAASADPRGRTEIKRLGSVTRVRSAAADARRVDDLRGFGA
GILTPSDTGGIALLLAHGRLEQARSVLEQGRPGDGDARTAAARQLAEIRDSVLNR
CAWCAGTRAAMHENPAGRPSTGRAGTGLRWLSGSHAALLAQRTAGTEGLLEITHTLD
ATFDVLNAVWTLALGAVERALHWCOTFLREAKERDAPGWEAVFAA VRAEGALRGGN
LAEARDARCFVLPERHSALSGSAVSILVTYQIARGKYEDARQOLSRPWADSLPD
TIHWLDYLRAGLYMATRQYHSALRDHPHEVRLAERWGTOWPALLPRTDTEALLR
LGREQRLGATELQAVQDGAPRIKGISLRQLGLASPEARLLNRAVLSADD
RLELSLTFDLADARETROQAMVLRARQLAKSCCAQPPDRIRISPDPAASGE
LFGCPERNADWGHLSERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
INGREELLAEARAGVQV"

7658..10486

/gene="olmRII"

7658..10486

/gene="olmRII"

/codon_start=1

/transl_table=11

/product="luxR-family transcriptional regulatory protein"

/protein_id="BAB69190.1"

/db_xref="GI:15823973"

/translation="MLASGQEGGVVWVWAGVGVGRTALLDAFLRKATSGARVCAA
AGSATEAEFGVWQLFPPTPLGAGFRLARIGSDAGPLGDALDFDLCOELRR
RPTVIADVDAQADAASIRFLRYLVRRLTAPVMVLTPTGPDALPSPFOALRHLP
QCRILGLTSDRGVTVQVSAQSDRVPRLAADPHALISGNGNPLLVALLHDLRAHR
LSPVSPVTVPSVSSAASAAASAPAAAPAAAFKAAALAWAYRDEPVLV
DVA CGFVLGESATPARVACLVDRGADAVQDMTALDTAGLLAGTFFSALVKGALL
HVDVETRAELERRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMOEGR


```

RQGGTLDFHLSLAQRHGHETSATTVLSARLTALSPTQQSLLDLVRAHMTAVMLND
DGNERTADGASFSFAHLGSPDSVMGVELNRSLKATGLRLVETLIFDHTPPAAVAAR
LRTAALHDEDTAPVDPSPSGHGTAAADPIAIIWACRFPGVGRSPKOLWELAAS
GGDAIGPPTDGMTEQORHAODPTQPGTFYQGGGFLHDAHFADAGPFGISPREALA
MDPQRLLETTSWEAFERAGIDPTLVGRSRTGVFAGALSFDGPRMDRTASSGGAADVE
GHILGTGSLVSGRAIYVFGLEGPAITVDTCGSASLVTLHIALCQSDRSTGCTALAG
RYSVMTLGMFTIEFQRGLSGLDGRCKAYSAADAGTGWGEVMLLVRLSDAVRLGH
GYLVNAGVAVNQDAGSNGLTAPNGPAQERVIRQALANAGLSADVAVVEHGCTGTL
GUP IEQAALATYQORAGRPLWGLSKNIGHTMAAGGVGVIKMMALRGVLPRT
LHVDPSPQDVASAGARLLTEAVPWGDAACRLBRAGVSFSGIGCTAHNVILLEAP
AGCVAGGCVLEGAGLAI SVAESVAA PVVASAPVAESVVPVVPVPSVARSSEAG
LRAQEAUKYVAVPVDVSLDVGAGLACGRVALEHRAVLAADREELVQGLGALAAP
EDPRVETGHPVGRDGRGVVFPFGCGAGMGVRLASSVPFARVQACBEALAPW
VQWVSVDLIRRDGAQWVERADVAQVPLFSVMVLSAALWSPISPEDAVLGSGOEIA
AAHVCGALSIDKAAKTVALRSALAAVGRGGMASVPLPAQEVOLIGERWAGRLWA
AVNGRPSFVSGDAEAVDEVLAYCAGTGVARERI PVDYASHCPHQVPLREELILIGD
ISFQSGVPFSTVEGTLDLTDLOAAWYRNHLGVRPFSADVQALADGHRVFEVS
PHPTLPAIATEDTEDTAEDVTAIGSLRGDNDTRPLTALAHHTTGITGPTWHHYH
THHHTPHPHTHLDTLPYFQHYWLESOPFGSGGAGSGAGSGAGSGRAGTAGCTA
EVESRFWADVQDLLETVA TLAVPSPAGLDTVPALSAWHRHQDQARINTWYQET
WKPLTLPTHQPHQTLWIAI PETQTHPHI TNILTNLHHGITIPLTNLHHTNPOH
LHHTLHTRQQAQNTTCATIGLLSILLADETPHHPHTPTGTLNLTLCQTHQTH
PTPLWYATNATTHNDPLTHPTQAQWGLARTILLEPHHTAGIIDLPTTPHT
LQHLTQLTQPHQQLAIRTTGTTRRLTPTLTPHTPTPTPTPTGTLITGGTGAL
ATHLTHLTHPTQHTQLLTSRTGHTPHQHLTTLQOQKQHLTITTCDSNPDQLQ
QLLNTIPPOHPLTVHTAGILDDATLNLPTQLNVNLRKAKHSAHLHQLTOHTPL
TAFVLYSSAATFGAGQANYAANA YLDALAHRRHTHLPA TSIAWGTWQGLADS
DKARERAVORPTPPELHKTIAQTSACQRAALBELVRDVAANRHADPRKAIAPDQS
FRALGDSLTAKATGLRPLVSLFVDPHTPAKLAVHLQNGDPRKAIAPDQS
AAVPAESVETPIAIVGMACRFPGVTSADDFWLLISEQDAIGFPTDRAWETIHA
YVDPDASQTCVTNRGGFLYDAGFDABFFGISPREALAMDPOQLLLETAWETIHA
GINPHTLHCTPTGVTNGODYALRVHNAQOSTDGFALTGTAGSVISRSISYTGFE
GPVSDVTSACSSVALHIALCQALRAGCSMALAGVTVMSPPGAFVFPFSQRTGLAD
GHCKAFSAADGTGMEGVMLLVERLSDAHRNGHVRVLA VGRSVAVNOGASNGTAP
NGSQOVRVQALANAGLSAGDVAEAGHTTGLDPIEAQALIAITGQDRAGGFL
WLGSPVNVHTQAAAGVAGVIMVMALRHGLPRTLHVDEP SHVDMSAGSAGVLLTE
TPVWPGELSRARAGVSGVSTNAVILLEAPADDPVGGPAGCEGDAVDEAAAG
SPGVWGLKRSQALRAQALHAHLTDHPLGLDADVGYTLA HARAVFOHRA TLIA
ADRDPTLQALALAGEPHPAVHSSAPGGTGTGAAGTAFICSGQGTQRPMAHGL
YTHVPFLQALNDICTHLDPLHDLPLTQNDNEDAAALQOQTRIAQALPAFQ
VALHRLTDGVIHTHYAGHSLGELTAHLAGILTDATLTITQATLITQATLPAFQ
LPTNHAFSPHTNPLNQLHQHTQTLTHPHPTPLITANTPDQLTTPHYWTOQART
VDYATOTLTHQGVTVYIELGPDNTLTLTHNLPNPTTLTTLTHPHHPTQHLT
NLAQTTTHPHHYTHDNOHPTHTLDTLPYFQHHYVLESTQPGAGNVAAGLDP
TEHPLGLALELATOGALLAGLSRHPMLDAHAGVLLSGTLELALHAGTY
VGCDDVDELTHAPLVVVDGVGVQVGAADGEGRLSVYARGGACGGGASGG
VMTCHASGVLVEAAGGVVDGLAGVMPRGA VAVDVGDRDLRAGACGLVPFVSGL
RAVRMDGDLALAEVCLPEAAGDAAGFGLHPALLDGVVQPLSVLLPGGTGEGAGFG
EGVRVPAVMGVSLHRAQVTVRVRVAVRGGGREA VSVVVDGAGVPAVSDRLEL
RPVDMQLRVAVSVAGRGLSYAVQAEVGVVPCQAWAHEDEVGESGGGPVGVV
LRCPDAGAGGGGGGGGGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV
DGPDVGVASVWGLVRSQAHPDRFVLLDLDTDGTDLDCAGAGWGDGAVAGPE
ACGEPLAVRGLLVAALKRLESQDPAQGRARRSDVPAGSGGVPARRSVD
VSGREPLWLSGSLVVTGTVGLGAARHLAGVGVDRDILLVSRGPDPAAGBLR
AELALGAERVIAQDVGEREVVELLEGVAGCPLCTGVVHAAGDADATIASLTPER
LGTVPAKVDALLDELITRAGMSAFVLFSSAAGILSGACGNYAANAALDALAY
KRAAGPGLWANGLESAGMTHGLDHRRIIRSGHPMSTPDALALFDALALYD
RPVLLPADLRAPPLPPLLOLLPATRRTTTTGGADNGALHARLAGTQHQH
TLLALVRSHATVLGHTTPTDIPDRAFDIGFDSLTAVELNRKLSLGTGLRPTTL
AFDHPNPTTLTHLHTQLOPDNAPA VLAELDKLESALDKTDKTSASERVTLRK
SMLKNWPAQHTAESADDEKFTSATAEIIFKFDINDLGLS"
12071..30790
/gene="avea2"
/contig="30790
/contig_start=1
/contig_end=11
/contig_table=11
/product="type I polyketide synthase AVES 2"

```

gene

CDS

```

/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MOLANEAKLLEYLRKRVADLDRTRRLYEVVEREOPETIAIVGMA
CRYPGATSPRLMHLVKSQTDAGEFPTDGRGNLEOLYDDPDPSYTSRSGSGFLY
DAGFDAAFPRLSPREALAMPQORLLLETTHETWETFEQGIIDPRSMRSGRTGVFVGNP
EDYTYTHQSPNAVGYLLTGSAAISAGRI SYNFEGLEGPAITITDTCSSLSVLAHL
ACQALRSECTMALAGASVMTAFVTFEQRQGLAADGRCKAFSAADAGTGWSEGV
GMLVRLSGDNRHGRVLA VGRSVAVNQDAGSNGLTAPNGRSQVAVIRQALANAGLS
PADVDAEAGHTTGLDPI IEQAALVEAYGQDRPNRPLWGLTGLTGNISGTHWAAAGV
GVIKMMALRNGLLPRTLHVDEPSPHVDMSAGVOLLTETVWPGCEGRLRAGVSVF
GVSGTNAHVILLEAPAHNIPSDTPADDPVGSADDEDAGSDEAAAGSPGVWPLVSA
KSOALAPQAQALHAHLTDHPLGLDADVGYTLA HARAVFOHRA TLIAADRTFLQALQ
PALAGEPHPAVHSSAPGGTGTGAAGTAFICSGQGTQRPMAHGLYTHVPFLQALNDI
NDICTHLDPLHDLPLTQPHQHTLTLLEAAALQOQTRIAQALPAFQVALLHRL
LTDGVIHTHYAGHSLGELTAHLAGILTDATLTITQATLITQATLITQATLITQATL
TPHHTLHHTAHENDLAAIANTPSLVSIGTPTVQHLTTLCOQGIKNTPLTNHA
FHSHTNPLNQLHQHTQTLTTHPHPTPLITANTPDQLTTPHYWTOQARTVDAAT
TOTLHQHGVTVYIELGPDNTLTLTHNLPNPTTLTTLTHPHHPTQHLTTLNLAAT
TTPHHPHTHNNQPHHTHLDLTPYFQHHYVLESTQPGSDPSSEGEQAOTPTST
PLRDLVKGSPQERDEELRVTHAAVLAHATPEVTPNKAFLKELAFDSLAAIQLR
NRLADVDLPATLIDYPTPMALCOFLRAAIVGADTGTTLTTLPLTAVPADEPIAIV
CMACRYPGDVTVDLQVGVSGHDAJGFFPNRSGWDLTLXNPDHGHGTSYTSRSG
FLYDAGNDPDPFGISPREALAMPQORLLLETWESIEHACINPDSLRGCTGTGVPAG
LTYHDAARFPTAPAGFEGYLGHSAGSIA SGRVAYALGLEGPALTVDTACSSSLVAL
HLAQAALRSECTMALAGVTVMSPTAGVFSRQGLAVDGRCKAFSAADAGTGWGE
GVGMLLVERLSADARLHRIILAVGRSVAVNQDAGSNGLTAPNGRSQVAVIRLANAD
LT PADVDAEAGHTTGLDPI IEQAALATYQDRPGNEPLWGLSKNINIGHAQAAG
VGGVIMVMALRNGLLPRTLHVDEPSPHVDMSAGVOLLTETVWPGCEGRLRAGVS
SFGVSNVAHVILLEAPAHNIPSDTPADDPVGAADDPVGAAGDGTGGEATGPA
AGSGVWPLVSAKSPALRAQALHAHLTDHPLGLDADVGYTLA HARAVFOHRA TL
IAADRTFLQALALAGEPHPAVHSSAPGGTGTGAAGTAFICSGQGTQRPMAHGL
GLYTHVPFLQALNDICTHLDPLHDLPLTQNDNTQTLTLEAAALQOQTRIAQAL
PALPAFQVALLHRLTDGVIHTHYAGHSLGELTAHLAGILTDATLTITQATLITQATL
QTMPEGTWTLTTPHHTLHHTAHENDLAAIANTPSLVSIGTPTVQHLTTLCOQ
QGIKNTPLTNHAFSPHTNPLNQLHQHTQTLTTHPHPTPLITANTPDQLTTPHYW
QOARTVDAATTOTLHQHGVTVYIELGPDNTLTLTHNLPNPTTLTTLTHPHH
PQHLTTLNLAATTTWHPHTHNNQPHHTHLDLTPYFQHHYVLESTQPGSAGS
QSGRAGTAGTAEVSRPMDAVARODLETVA TLAVPPSAGDITVVPALSAWHRH
QHQDARINTWYQETWKPILTTPHQTQHLTSRTGHTPHQHLTTLTTPHTPTPTAGI
TCLTGTGALATHLTHHTHQTQHLTSRTGHTPHQHLTTLTTPHTPTPTAGI
TCDTSNPDQLQNTIPPOHPLTVHTAGILDDATLNLPTQLNVNLRKAKHSAH
LHQLTOHTPLNAPVLYSSAATFGAGQANYAANA YLDALAHRRHTHLPA TSIAW
GTWQGLADTQVSEHRRRCHWFAMPPELATVNDGAIASGRPSLLVADI DKKLGPV
LSSSVLLDPLQOQGTTEARSTVEQTESLNRKQLMGRSRESEELSLVLRTHSA
AVLGRDSEALPPGRLFRDLGSLAAVELNRHLLAAQTELALPTLTVDFSPSTKLQAI
FLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGVTSADDFWLLISEQDAI
GGFTDRGMDLTDLYDDPDHPGTCYTRNGGFLYDAGFDABFFGISPREALAMDPOQ
RLLETAWETIEHAGINPHTLHGTPTGVTNGODHAAHRIQAPSGTGVLTGAAT
SIASGRISYILGLEGPALTDLTACSSSLVALHIALCQSLRSGECTMALAGTAVMTPI
TFTBPAQRGLADGRCKAFSAADAGTGWGMLLVERLSDAHRNGHVRVLA VGRSV
AVNQDAGSNGLTAPNGRSQVAVIRQALANADLT PADVDAEAGHTTGLDPI IEQA
LATYQDRPGNPLWGLSKNVGHTQAAAGVAGVIMVMALRHRLTLPRTLHVPDGP
HVDASAGVOLLTETVWPGCEGRLRAGVSVFVSGTNAHVILLEAPADDPVGGP
DEDAGSEEAAGSPGVWPLVSAKSPALRAQALHAHLTDHPLGLDADVGYTLA
ARA VDRHRA TLIAADRTFLQALALAGEPHPAVHSSAPGGTGTGAAGTAFIC
QGTQPGMAHGLYTHVPFLQALNDICTHLDPLHDLPLTQNDNEDAAALQOQTRIA
LOQTPYAQALPAFQVALLHRLTDGVIHTHYAGHSLGELTAHLAGILTDATLT
ITQATLITQATLITQATLITQATLITQATLITQATLITQATLITQATLITQATL
HITTLCOQGIKNTPLTNHAFSPHTNPLNQLHQHTQTLTTHPHPTPLITANTPD
OLLTPHYWTOQARTVDAATTOTLHQHGVTVYIELGPDNTLTLTHNLPNPTTL
TLTHPHHPTQHLTTLNLAATTTWHPHTHNNQPHHTHLDLTPYFQHHYVLE
PSAQSPQRSSRSAPDAESEFWDVANEEDLSLAETLIDISALDTVVPALSAWH
RHQDQARINTWYQETWKPILTTPHQTQHLTIAI PETQTHPHI TNILTNLHH
ITPILTVNHTNPOHLLHTRQQAQNTTGTITGLLSALADSTPHPHPTPT
TGTLLNTLTPOTHTPTPLWYATNATTHNDPLTHPTQAQWGLARTILLEPH
THTAGIIDLPTTPHTLHHTLTHPTQOLAIRTTGTTRRLTPTLTPHTPTPT
TPTPTGTLITGGTGALATHLTHHTHQTQHLTSRTGHTPHQHLTTLTQOQK
IHLTITCDSNPDQLQNTIPPOHPLTVHTAGVNLVPAFVSETDAESFSSVTA
KATGAALHLLHDLHETLIFSSGAGAWGSCNQCA SAANYLALAHRTQHTGL
PGASIAWPGWAGKGSADAAHGYLEKRGILPMEPRMALAAAFHRAQRPNSNLIAD

```

Patent: WO 0001827-A 1 13-JAN-2000;
KEILLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PRIZER (US); MCARTHUR
HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)

FEATURES

BASE COUNT	1884 a	4561 c	4005 g	1931 e
ORIGIN				

**BASE CO
ORIGIN**

Query Match	100.0%	Score 258;	DB 6;	Length 12381;
Best Local Similarity	100.0%;	Pred. No. 2e-32;		
Matches 258;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CTGACCGCGTGTCTCCACGACGACGAGTCTGCTCTCTGGACCTGGTTCGGCGCCAC	60	
Db	1913	CTGACCGCGTGTCTCCACGACGACGAGTCTGCTCTCTGGACCTGGTTCGGCGCCAC	1972	
Qy	61	ACCATGCGCGTGTCTGAACGACGACGGAACGAGCGCACCGCGTTCGGATCGCGCGCCATCG	120	
Db	1973	ACCATGCGCGTGTCTGAACGACGACGGAACGAGCGCACCGCGTTCGGATCGCGCGCCATCG	2032	
Qy	121	GCGAGTTTCGCCACACCTTCGGCTTCGACTCCGTCATTCGGGTTCGAACTGCGCAACCGCGCTC	180	
Db	2033	GCGAGTTTCGCCACACCTTCGGCTTCGACTCCGTCATTCGGGTTCGAACTGCGCAACCGCGCTC	2092	
Qy	181	AGCAGGCGACGGGCGTTCGGTTCGCCCGTGACGCTCATCTTCGACACACACGACGCGGCC	240	
Db	2093	AGCAGGCGACGGGCGTTCGGTTCGCCCGTGACGCTCATCTTCGACACACACGACGCGGCC	2152	
Qy	241	GCGGTTCGCCCGCGCGCTT	258	
Db	2153	GCGGTTCGCCCGCGCGCTT	2170	

RESULT 2					
E38020					
LOCUS	E38020	Avermectin aglycon synthase gene.	30690 bp	DNA	linear
DEFINITION	E38020				
ACCESSION	E38020				
VERSION	E38020.1	GI:18626909			
KEYWORDS	JP 2000245457-A/1.				
SOURCE	Streptomyces avermitilis.				
ORGANISM	Streptomyces avermitilis				
	Bacteria; Actinobacteria;				
	Actinobacteridae; Actinomycetales;				
	Streptomycineae; Streptomyces				
	Streptomycetaceae; Streptomyces.				

REFERENCE	Streptomyces
AUTHORS	Omura, S. and Ikeda, H.
TITLE	Avermectin aglycon synthase gene
JOURNAL	Patent: JP 2000245457-A 1 12-SEP-2000;
	THE KITASATO INSTITUTE
COMMENT	OS Streptomyces avermitilis
	PN JP 2000245457-A/1
	PD 12-SEP-2000
	PF 24-FEB-1999 JP 1999046961
	PR
	PI SATOSHI OMURA, HARUO IKEDA
	PC C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
	C12Q1/68, C07D493/22,
	PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), C12N15/00 CC
	Location/Qualifiers
FH Key	(1) . . (11916)
FT CDS	(11971) . . (30687)
FT CDS	

```

FEATURES
source
Location/Qualifiers
1. 30690
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 5355 a 12455 c 8617 g 4363 t
ORIGIN

```

ORIGIN

Query Match 100.0%; Score 258; DB 6; Length 30690;
Best Local Similarity 100.0%; Pred. No. 1.6e-32;

	Matches	258;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	CTGACGCGTGTC	CCCCACG	CAGCAGC	AGT	CGCTCT	CTGTG	ACCTGTG	CGCGCCAC	60
Db	1096	CTGACGCGTGTC	CCCCACG	CAGCAGC	AGT	CGCTCT	CTGTG	ACCTGTG	CGCGCCAC	1155
Qy	61	ACCATGCGGTGCT	GAAACG	ACGAGC	AGAA	CGACG	CAC	CGCTG	CGATCGCGCCCATCG	120
Db	1156	ACCATGCGGTGCT	GAAACG	ACGAGC	AGAA	CGACG	CAC	CGCTG	CGATCGCGCCCATCG	1215
Qy	121	GCAGGTTTCC	CCACCT	CGCGTT	TCG	ACTCG	CTCAT	TGGTGT	CGAACTGCGCAACCGCCTC	180
Db	1216	GCAGGTTTCC	CCACCT	CGCGTT	TCG	ACTCG	CTCAT	TGGTGT	CGAACTGCGCAACCGCCTC	1275
Qy	181	AGCAAGCCAC	CGGCGT	GGGTTG	CCCGT	GACG	CTCAT	TTTCG	ACACACACCGCGGGCC	240
Db	1276	AGCAAGCCAC	GGGCGT	GGGTTG	CCCGT	GACG	CTCAT	TTTCG	ACACACACCGCGGGCC	1335
Qy	241	GCGGTCGCG	CGCGCCTT	258						
Db	1336	GCGGTCGCG	CGCGCCTT	1353						

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 64957)
Ikeda, H., Nonomiya, T., Ueami, M., Ohta, T. and Omura, S.
Organization of the biosynthetic gene cluster for the polyketide
anthelmintic macroide avermectin in Streptomyces avermitilis
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
99380548

2 (bases 1 to 64957)
Ikeda, H., Nonomiya, T., Ueami, M., Ohta, T. and Omura, S.
Direct Submission
Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
Sciences, Kitasato University, Microbial chemistry; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel. +81-3-5791-6242,
Fax. +81-3-1444-6197)

```

FEATURES
source
    1. .64957
        Location/Qualifiers
            /organism="Streptomyces avermitilis"
            /db_xref="taxon:33903"
            101..12019
                /gene="aveA1"
            101..12019
                /gene="aveA1"
            /notes="contains a loading module, Ave module 1 and module
            2
            multifunctional polyketide synthase"

```

```

/transl: "cabresil"
/product: "type i polyketide synthase AVES 1"
/protein_id: "BAA8474.1"
/db_xref: "GI:5902891"
/translation: "MQMDGGEPRPAAGEVLGVADQGVVVFPGQWPGMGR
ELLDSDFVRESVRACEAFAPYVDMSQVLRSDSPDFCLDRVDVVQPTLFAVMSLR
AALLRQSGVEPCAPLWGHSGLEIAAHVSGGLSLADAAVVYTLWMSQAQTTLAGTAVR
VAATPDELLPAPTWEDENPARLAAVANGSPRSTVSGSAREAVADVLADTAAGVTR
MIPVDVPSAPLMLYAEERVVSPGLPITPRPSIIPFHSSVTCGRRLDRELDAAYVYRN
MSSTVFAPEARLLQDQGPKTFFVMSPLVLTMLGLOELPDLDGDTTGTADTVMGLTR

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 500.468 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353

Perfect score: 258
Sequence: 1 ctgacgcgctgtcccccac.....ccgcggtcgccgcgcctt 258

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	258	100.0	12381	6	AX006889 Sequence
2	258	100.0	30690	6	E38020 Avermectin
3	258	100.0	64957	1	AB032367 Streptomy
4	256.4	99.4	11096	1	AF275943 Streptomy
5	95	36.8	104326	1	AB070940 Streptomy
6	93.4	36.2	123580	1	AF263912 Streptomy
7	93.4	36.2	125401	6	AX211739 Sequence
8	91.8	35.6	43280	1	SFU78289
9	91.8	35.6	104326	1	AB070940 Streptomy
10	90.2	35.0	53784	1	AMM23012
11	90.2	35.0	53789	6	A69720 Sequence 3
12	90.2	35.0	90445	1	AF040570
13	87	33.7	30000	6	AX250263 Sequence
14	85.4	33.1	5676	6	A69718 Sequence 1
15	85	32.9	78210	1	AB070949 Streptomy
16	84.6	32.8	32870	1	AF007101 Streptomy
17	83.8	32.5	20394	1	SNA13222
18	83.8	32.5	20394	6	AX067996 Sequence
19	83.8	32.5	84985	1	SNA278573
20	83.2	32.2	77534	1	AF235504
21	82.6	32.0	2700	6	AX211733 Sequence
22	82.6	32.0	2706	1	AF420594 Streptomy
23	82.6	32.0	65140	6	AX211705 Sequence
24	82.4	31.9	5544	1	SNA132221
25	82.4	31.9	5544	6	AX067994 Sequence
26	82.4	31.9	84985	1	SNA278573
27	82.4	31.9	113193	1	AF357202
28	81.6	31.6	30000	6	AX250263 Sequence
29	80.6	31.2	41097	1	AF016585
30	79.8	30.9	82746	1	AF453501 Actinosyn
31	79.4	30.8	37948	1	AF079138 Streptomy
32	79	30.6	24225	1	SC2C4
33	79	30.6	28732	1	AF082100 Streptomy
34	77.6	30.1	26195	1	SC1G7
35	77.4	30.0	1681	6	AR173228
36	77.4	30.0	11219	1	SERERYAA
37	77.4	30.0	11219	6	AR049367
38	77.4	30.0	11219	6	AR095528 Sequence
39	77.4	30.0	31422	6	E38021
40	77.4	30.0	64957	1	AB032367 Streptomy
41	75	29.1	758	1	AF144051 Streptomy
42	74.6	28.9	27522	1	AB070942 Streptomy
43	74.6	28.9	30000	6	AX250262 Sequence
44	74.4	28.8	107379	1	SHGCCPIR
45	74.2	28.8	107379	1	SHGCCPIR

ALIGNMENTS

RESULT 1
AX006889
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AX006889
Sequence 1 from Patent WO0001827.
AX006889
AX006889.1 GI:9994904

Streptomyces avermitilis.
Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

REFERENCE
AUTHORS
Kellenberger, J.L., Leaday, P.F., Staunton, J., McArthur, H.A. and
Stutzman-Engwall, K.J.

Polyketides, their preparation, and materials for use therein

PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

PA (ALPH-) ALPHARMA AS.

PA (SINV-) SINVENT AS.

PA (DZIE/) DZIEGLEWSKA H.

PA (ZOTC/) ZOTCHEV S B.

PA (SEKU/) SEKUROVA O N.

PA (FJAE/) FJAERVIK E.

PA (BRAU/) BRAUTASET T.

PA (STRO/) STROM A R.

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

DR P-PSDB; AAE10153.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals -

XX Example 2; Page 186-187; 266pp; English.

XX The present invention relates to the cloning and sequencing of the gene

CC cluster encoding a modular type I polyketide synthase (PKS) enzyme

CC involved in the biosynthesis of the macrolide antibiotic nystatin.

CC The nystatin PKS is useful as antifungal antibiotics. The present

CC sequence is a BRD48 insert DNA in pKR48 which is used in an

CC exemplification of the invention.

XX Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;

SQ Query Match 32.0%; Score 82.6; DB 22; Length 2700;

Best Local Similarity 57.6%; Pred. No. 7.2e-09;

Matches 148; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1 CTGACGCGGTGTCGAGACGACGAGGAAACGAGCGACCGCGTCGGATCGCGCCCATCG 120

Db 1962 CTCGCGGACGAGTCACCGGCTCCAGCGCGCGGACGAGGACACCTCTCTCGGCTC 2021

QY 61 ACCATGCGGTGTCGAGACGACGAGGAAACGAGCGACCGCGTCGGATCGCGCCCATCG 120

Db 2022 GTCCGCGACGACGCGGATGTCCTCGGCGACACCTCGGCGGACGCGGTGACCGCTCC 2081

QY 121 GCGAGTTTCGCCACCTCGGCTTCGACTCGGTATGGGTGTGCACTGGCGCAACCGCTTC 180

Db 2082 CGGCGCTTCGCGACCTCGGCTTCGACTCGCTCAGCGGTCGAACTCGCGCAACCGCATC 2141

QY 181 AGCAAGGACGCGGCTCGGCTTCGCGTACGCTCATCTTCGACACACGACGCGCGCC 240

Db 2142 GCGCGGCGCACCGGCTCGGCTTACCGCGGACGCGGCTCTTCGACTACCGCCCGCGAT 2201

QY 241 GCGGTGCGCGCGGCTCT 257

Db 2202 GCCTCGCGCACACT 2218

RESULT 12

AD17184

ID AD17184 standard; DNA; 65140 BP.

XX AD17184;

AC AD17184;

XX 29-NOV-2001 (first entry)

DT Streptomyces noursei nysI DNA of nystatin PKS gene cluster.

DE Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX antifungal; antibiotic; nysI; ds.

KW Streptomyces noursei.

XX Location/Qualifiers

OS complement (1..1035)

PH *tag= a

FT

FT /product= "NysD2 partial protein"

FT /note= "CDS does not include stop codon"

FT complement (1056..2576)

FT *tag= b

FT /product= "NysD1 protein"

FT 2806..6906

FT /product= c

FT /product= "NysA protein"

FT 6952..16530

FT *tag= d

FT /product= "NysB protein"

FT 16550..49840

FT /tag= e

FT /product= "NysC protein"

FT 50260..51015

FT *tag= f

FT /product= "NysE protein"

FT 51405..54305

FT /tag= g

FT /product= "NysR1 protein"

FT 54329..57190

FT *tag= h

FT /product= "NysR2 protein"

FT /note= "CDS does not include start codon"

FT 57180..59963

FT *tag= i

FT /product= "NysR3 protein"

FT 60415..61047

FT *tag= j

FT /product= "NysR4 (short) protein"

FT /note= "CDS does not include start codon"

FT 61736..62497

FT *tag= k

FT /product= "NysR5 protein"

FT /note= "CDS does not include start codon"

FT complement (62551..63615)

FT *tag= l

FT /product= "ORF2 protein"

FT /note= "CDS does not include start codon"

FT 63765..64961

FT *tag= m

FT /product= "ORF1 protein"

FT WO200159126-A2.

FT 16-AUG-2001.

XX 08-FEB-2001; 2001WO-GB00509.

XX 08-FEB-2000; 2000GB-0002840.

XX 10-APR-2000; 2000GB-0008786.

XX 14-APR-2000; 2000GB-0009387.

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.

XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

XX (ALPH-) ALPHARMA AS.

XX (SINV-) SINVENT AS.

XX (DZIE/) DZIEGLEWSKA H.

XX (ZOTC/) ZOTCHEV S B.

XX (SEKU/) SEKUROVA O N.

XX (FJAE/) FJAERVIK E.

XX (BRAU/) BRAUTASET T.

XX (STRO/) STROM A R.

XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;

PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;

XX WPI; 2001-557614/62.

DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,

DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.

XX New nystatin polyketide synthase polynucleotides and polypeptides,

PT useful as antibiotics and antifungals -

```
FT      /note= "KS4"  
FT      complement (28750..28960)  
FT      /tag= ad  
FT      /note= "ACP3"  
FT      complement (29092..29740)  
FT      /tag= ae  
FT      /note= "KR3"  
FT      complement (29869..31018)  
FT      /tag= af  
FT      /note= "DH3 (inactive)"  
FT      complement (31018..32185)  
FT      /tag= ag  
FT      /note= "AT3"  
FT      complement (32185..33439)  
FT      /tag= ah  
FT      /note= "KS3"  
FT      complement (33505..33715)  
FT      /tag= ai  
FT      /note= "ACP2"  
FT      complement (33823..34480)  
FT      /tag= aj  
FT      /note= "KR2"  
FT      complement (34606..35749)  
FT      /tag= ak  
FT      complement (28750..28960)  
FT      /tag= al  
FT      /note= "DH2 (inactive)"  
FT      complement (35749..37144)  
FT      /tag= am  
FT      /note= "AT2"  
FT      complement (37145..38296)  
FT      /tag= an  
FT      /note= "KS2"  
FT      complement (38371..38581)  
FT      /tag= ao  
FT      /note= "ACP1"  
FT      complement (38677..39307)  
FT      /tag= ap  
FT      /note= "KR1"  
FT      complement (39442..40609)  
FT      /tag= aq  
FT      /note= "DH1"  
FT      complement (40609..41842)  
FT      /tag= ar  
FT      /note= "AT1"  
FT      complement (41842..43093)  
FT      /tag= as  
FT      /note= "KS of extender module 1 (KS1)"  
FT      complement (43144..43660)  
FT      /tag= at  
FT      /note= "ACP of loading domain"  
FT      complement (43777..44829)  
FT      /tag= au  
FT      /note= "ER of loading domain"  
FT      complement (44974..46573)  
FT      /tag= av  
FT      /note= "CoA ligase of laoding domain"  
FT      46754..47788  
FT      /tag= aw  
FT      /note= "fkbo gene"  
FT      47785..52272  
FT      /tag= ax  
FT      /note= "fkbp gene"  
FT      52275..71465  
FT      /tag= ay  
FT      /note= "fkba gene"  
FT      52362..53576  
FT      /tag= az  
FT      /note= "KS7"  
FT      53577..54716  
FT      /tag= bb  
FT      /note= "AT7"  
FT      54717..55871
```

```
FT      /tag= ba  
FT      /note= "DH7"  
FT      56019..56819  
FT      /tag= bc  
FT      /note= "ER7"  
FT      56943..57575  
FT      /tag= bd  
FT      /note= "KR7"  
FT      57710..57920  
FT      /tag= be  
FT      /note= "ACP7"  
FT      57990..59243  
FT      /tag= bf  
FT      /note= "KS8"  
FT      59244..60398  
FT      /tag= bg  
FT      /note= "AT8"  
FT      60399..61412  
FT      /tag= bh  
FT      /note= "DH8 (inactive)".  
  
Query Match      32.2%; Score 83.2; DB 21; Length 77536;  
Best Local Similarity 62.5%; Pred. No. 5.2e-09;  
Matches 130; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
  
QY      50 TCGCGCCACACCATGCGGTGCTGAACGACGACGCGAAACGAGCGCGCGTCCGATG 109  
Db      38580 TGCTGCGCCATCGTGTGTGCGGCCACGCGCGCGCTCGGCCACGCCGCTCCGAGA 38521  
  
QY      110 CCGGCCCATCGCGAGTTTCGCCACCTCGGCTTCGACTCCGCTATGGGTGTCGAATGC 169  
Db      38520 TCACGCCACGACGCGCTTCAAGGACCTCGGACATCGCTCAGCGGTGTCGCGTTGC 38461  
  
QY      170 GCACCGGCTCAGCAGCGCCACGCGCTCGGTTGCCGTCAGCTCATCTTCGACCACA 229  
Db      38460 GCACAGGCTCGCCGAGACGACGCGGGTACGCTCTCCGCGACGCGCGTCTTCGACCACC 38401  
  
QY      230 CCACCGCGCGCGGTGCTCGCGCGCGCT 257  
Db      38400 CGACACCGCGCGGTGCTCGCGCGCGCT 38373  
  
RESULT 11  
AADI7193  
ID AADI7193 standard; DNA; 2700 BP.  
XX AC AADI7193;  
XX DT 29-NOV-2001 (first entry)  
XX DE ERD48 insert DNA in pKR48.  
XX KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;  
XX KW antifungal; antibiotic; ds.  
XX OS Streptomyces noursei.  
XX FH Key Location/Qualifiers  
FT CDS 3..2699  
FT FT /tag= a  
FT FT /product= "Protein encoded by ERD48 insert DNA"  
XX PN WO200159126-A2.  
XX PD 16-AUG-2001.  
XX PF 08-FEB-2001; 2001WO-GB00509.  
XX PF 08-FEB-2000; 2000GB-0002840.  
XX PR 10-APR-2000; 2000GB-0008786.  
XX PR 14-APR-2000; 2000GB-0009387.  
XX PA (UYN0-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
```

CC	The present sequence encodes a polyketide synthase which is associated	
CC	with the biosynthesis of pimarinin. The polyketide synthase polypeptide	
CC	is useful for the oxidative modification of a methyl group of a suitable	
CC	compound, e.g. a bioactive compound including a secondary metabolite,	
CC	antibiotics and anticancer agents. Recombinant cells comprising the	
CC	gene are useful for the production of pimarinin. The polyketide synthase	
CC	polynucleotide may be over expressed in Streptomyces, leading to an	
CC	increase in the biosynthesis of pimarinin, as a source of primers for	
CC	amplification reaction and as probes.	
XX		
SQ	Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;	
	Query Match 32.5%; Score 83.8; DB 22; Length 20394;	
	Best Local Similarity 61.5%; Pred. No. 4e-09;	
	Matches 158; Conservative 0; Mismatches 87; Indels 12; Gaps 1;	
Qy	1 CTGACCGCGTGTCCCCACGACGACGAGTCGCTGCTCTCTGGACCTGGTCGGGGCCAC 60	
Db	9574 CTGAGCGGACGGGAACGGTGGAGCGGCGGAGGTGCTCTCTGGACCTGGTATCGGGCCCGAG 9633	
Qy	61 ACCATGGCGGTGCTGAACGACGACGAGAAACGAGGCGCACCGCGTTCGATGCCGGCCCATCG 120	
Db	9634 ATCCGGTCTGCTCTGGGCGACGCGAACCAGGAGCATGAGTCCACCCGTGTC----- 9687	
Qy	121 GCGAGTTTCGCCCACTCGGCTTCGACTTCGCTCATGGGTGTCGAACTCGCAACCGCCTC 180	
Db	9688 -----TTCCAGGACCTCGGCTTCGACTTCCTGACCCGGGTGGAATCCGCAACCGCCTC 9741	
Qy	181 AGCAAGGCCACGGCGCTCGGTTGCGCGTGAAGCTCATTCCTTCGACCACACCGCGGCC 240	
Db	9742 AACAAACGCGACCGCGCTTCGCGCTTTTCGGCCACCGCGCTTCCTGACTACCCACGCGGAC 9801	
Qy	241 GCGGTGCGCGCGCGCT 257	
Db	9802 GCGCTGTCGACTTCCT 9818	
RESULT 10		
AAAL14651/c		
ID	AAAL14651 standard; DNA; 77536 BP.	
XX		
AC	AAAL14651;	
XX		
DT	08-AUG-2000 (first entry)	
XX		
DE	Nucleotide sequence of the FK-520 biosynthetic gene cluster.	
XX		
KW	FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;	
KW	Streptomyces hygroscopicus var. ascomyceticus; immunophilin;	
KW	FK-506 binding protein; polyketide compound; transplant rejection;	
KW	graft-versus-host disease; uveitis; alopecia universalis;	
KW	autoimmune chronic active hepatitis; inflammatory bowel disease;	
KW	multiple sclerosis; primary biliary cirrhosis; scleroderma;	
KW	neurite outgrowth; nerve regrowth; Parkinson's disease;	
KW	Alzheimer's disease; stroke; traumatic spinal cord; brain injury;	
KW	peripheral neuropathy; ss.	
XX		
OS	Streptomyces hygroscopicus.	
XX		
PH	Location/Qualifiers	
FT	complement (412..1836)	
FT	/tag= a	
FT	/note= "fkbw gene"	
FT	complement (2020..3579)	
FT	/tag= b	
FT	/note= "fkbv gene"	
FT	3969..4496	
FT	/tag= c	
FT	/note= "fkbr2 gene"	
FT	complement (4595..5488)	
FT	/tag= d	
FT	/note= "fkbr1 gene"	
FT	5601..6818	
FT	CDS	

Qy	61	ACCATGCGGTGTCGTAACGACGACGCGGTGCGATGTCGGGCCCATCG	120
Db	25291	GTCCGGCCGTGTCGCGGACGCGCA-----CGCAGAGCGCGTCGCGTCGAC	25338
Qy	121	GCAGATTTGCCCACTCGCTTCGACTCCGTCATGGGTGTCGAATCGCGCAACCGCTC	180
Db	25339	ACGGCTTTAAGACCGCGCTTCGACTCGCTGACCGCGGTGAGCTGCGCAACCGCATG	25398
Qy	181	AGCAAGGCCACGGGCGCTCGCGTTGCCCGTGAAGCTCATCTTCGACACACACCGCGGCC	240
Db	25399	CGGACCGCCACCGGCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGAACCCCGAG	25458
Qy	241	CGCGTGC CGCGCGCT	257
Db	25459	CGCTCGCCGGCACCT	25475

RESULT 8	
AAV21186	
ID	AAV21186 standard; DNA; 5676 BP.
XX	
XX	
AC	AAV21186;
XX	
XX	
DT	24-JUL-1998 (first entry)
XX	
XX	
DE	Amyncolatopsis mediterranei strain wT3136 5.7 kb KpnI DNA fragment.

AA	Key	Location/Qualifiers
PH	CDS	3..5676
FT		/tag= a
FT		/product= "polyketide synthase"
FT		/note= "no stop codon given"

XX	
FN	WO9807868-A1.
XX	
PD	26-FEB-1998.
XX	
Pf	18-AUG-1997; 97WO-EP04495.
XX	
PR	20-AUG-1996; 96EP-0810551.
XX	
PA	(NOVS) NOVARTIS AG.
XX	
PI	Engel N, Schupp T, Toupet C;
XX	
DR	WPI; 1998-169172/15.
DR	P-PSDB; AAMS2844.
XX	
PT	Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT	to produce rifamycin and rifamycin analogues
XX	
PS	Claim 4; Page 38-43; 205pp; English.

CC The present sequence represents Amycolatopsis mediterranei strain wt3136
CC 5.7 kb KpnI fragment DNA, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The

CC	DNA fragment can be used for assembling a library of polyketide
CC	synthases, which can be used for assembling a library of polyketides.
CC	A hybridisation probe of the invention can be used for identifying DNA
CC	fragments involved in the biosynthesis of ansamycins.
XX	
QQ	Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;
Query Match	33.1%; Score 85.4; DB 19; Length 5676;
Best Local Similarity	61.9%; Pred. NO. 1.8e-09;
Matches 159; Conservative	0; Mismatches 86; Indels 12; Gaps 1;
Qy	1 CTGACCGCGTGTCCCCACGACGACGAGCGAGTCTGCTCTCTGACCTGTGTGCGGCCAC 60
Db	3072 CTCGCCGGTCTCGCGCGCGGAGAGGAGCGCTGCTGCTCGACGTGCTCCGACGCGAG 3131
Qy	61 ACCATGCGGTGTGTAACGACGACGGAACAGAGCGACCGCGTCGGATGCGGCGCCATCG 120
Db	3132 GTGCGGTGTGTCTGGGACGCGCGGCGGAGCGCTCGCGGGACACGCGG----- 3185
Qy	121 GCGAGTTTCCCCACCTTCGCTTCGACTCCGTCATGGGTTCGAACTGCGCAACCGCCTC 180
Db	3186 -----TTCAAGGACACCGGCTTCGACTCGCTGCGTGGAACTGCGCAACCGGCTG 3239
Qy	181 AGCAAGCCACGGGCTGCGGTTGCCGTGACCGCTCATCTTCGACCACACCGCGCGCC 240
Db	3240 CGCGAGGCGAGCGGCTGAAGCTGCCGCGACGCTGCTCTTCGACTACCGACGCGGTC 3299
Qy	241 GCGTTCGCGCGCGCCT 257
Db	3300 GCGTTCGCGCGCTACCT 3316
RESULT 9	
AAF24892	
ID	AAF24892 standard; DNA; 20394 BP.
XX	
AC	AAF24892;
XX	
DT	20-APR-2001 (first entry)
XX	
DE	Pimaricin biosynthesis associated polyketide synthase gene.
XX	
KW	Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW	anticancer; pimaricin; ss.
XX	
OS	Streptomyces natalensis.
XX	
FH	Key Location/Qualifiers
FT	CDS 1..20394
FT	/*tag= a
FT	/product= "polyketide synthase"
XX	
PN	WO200077222-A1.
XX	
PD	21-DEC-2000.
XX	
PF	14-JUN-2000; 2000WO-EP06227.
XX	
PR	14-JUN-1999; 99EP-0201893.
XX	
PA	(STAM) DSM NV.
XX	
PI	Martin JF, Aparicio JF, Colina AJ;
XX	
DR	WPI; 2001-080693/09.
DR	P-PSDB; AAB31558.
XX	
PT	New polynucleotides encoding enzymes involved in the biosynthesis of
PT	pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT	the synthesis of new compounds
XX	
PS	Disclosure; Page 53-80; 116pp; English.
XX	

XX		27-AUG-1997.	
PD			
XX			
PF		19-FEB-1997; 97EP-0301056.	
XX			
PR		22-FEB-1996; 96US-0012078.	
XX		(ELIL) LILLY & CO ELI.	
FA			
PI		Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;	
XX			
DR		WPI; 1997-418046/39.	
DR		P-P5DB; AAW22601-W22605.	
XX			
PT		DNA encoding Streptomyces fradiae tyactone synthase domain - for	
PT		production of tylosin-related polyketide compounds	
XX			
PS		Claim 2; Pages 8-66; 220pp; English.	
XX			
CC		This sequence represents the tyactone synthase gene cluster of the	
CC		invention. This sequence is also referred to as the tylG gene, and was	
CC		isolated from Streptomyces fradiae. This sequence encodes multifunctional	
CC		proteins which direct the synthesis of the polyketide tyactone, isolated	
CC		from Streptomyces fradiae. Tyactone is the basic building block of the	
CC		antibiotic tylosin. The DNA sequence can be modified so as to alter the	
CC		type of carboxylic acids incorporated, the number of carboxylic acids	
CC		incorporated and/or the post-condensation reactions performed, thereby	
CC		resulting in novel tylosin-related polyketides.	
XX			
SQ		Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;	
		Query Match 35.6%; Score 91.8; DB 18; Length 43280;	
		Best Local Similarity 63.4%; Pred. No. 8.1e-11;	
		Matches 163; Conservative 0; Mismatches 82; Indels 12; Gaps 1;	
Qy	1	CTGACCGGTGTCGCCACGAGCAGCACTGCCTCTCTTGAGACTGGTGCGCCGCCAC 60	
Db	30648	CTGACCGGGTCTACGGCACAGGAAGGCGCCCTCTCTGCGCAGATCGCGCCCCAC 30707	
Qy	61	ACCATGCGGTGTGAACGACGACGGAACGAGCGCACCGCTCGGATCGCGGCCCATCG 120	
Db	30708	GCGGCGCGGTGTGGGCCACCGCTCCGACGACTCGATCCCAGGAGACGGGCC----- 30761	
Qy	121	GCGAGTTTCGCCACCTCGGCTTCGACTCGCTCATGGGTGTGAACTGCGCAACCGCCTC 180	
Db	30762	-----TTCAAGAACCTCGCTTCGACTCGCTCACGCCGTGAGATGCGCAACCGCTG 30815	
Qy	181	AGCAAAGCCACAGGGCTCGGTTGCCCGTAGCTCATCTTCGACCAACAACCGCGGCC 240	
Db	30816	AGCGCGGCCACCGGCTCTCGGCTGCCCGCACCTCTTCGACCAACCGCGGCCG 30875	
Qy	241	GCGGTGCGCGCGCCT 257	
Db	30876	GAGTGGCGGCCACCT 30892	
RESULT 7			
AAV21187			
ID	AAV21187	standard; DNA; 53789 BP.	
XX			
AC	AAV21187;		
XX			
DT	24-JUL-1998	(first entry)	
XX			
DE	Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.		
XX			
KW	Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;		
KW	polyketide synthase; actinomycete; ansamycin; ds.		
XX			
OS	Amycolatopsis mediterranei.		
XX			
Key	Location/Qualifiers		
FT	CDS	1825..15543	

RESULT 5

AAD17186
 ID AAD17186 standard; DNA; 125401 BP.
 AC AAD17186;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 KW antifungal; antibiotic; ds.
 XX
 OS Streptomyces noursei.
 XX
 PH Key Location/Qualifiers
 FT CDS 6337..34771
 FT /tag= a
 FT /product= "NysI complete protein"
 FT CDS 34792..51099
 FT /tag= b
 FT /product= "NysJ protein"
 FT CDS 51155..57355
 FT /tag= c
 FT /product= "NysK protein"
 FT CDS 57503..58687
 FT /tag= d
 FT /product= "NysL protein"
 FT CDS complement (58786..58980)
 FT /tag= e
 FT /product= "NysM protein"
 FT /notes= "CDS does not include start codon"
 FT CDS complement (59045..60241)
 FT /tag= f
 FT /product= "NysN protein"
 FT /notes= "CDS does not include start codon"
 FT CDS complement (60238..61296)
 FT /tag= g
 FT /product= "NysD2 complete protein"
 FT CDS 120628..121308
 FT /tag= h
 FT /product= "NysR4 (long) protein"
 XX
 WO200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-GB00509.
 XX
 PR 08-FEB-2000; 2000GB-0002840.
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.
 XX
 (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNTF) SINTEF STIPTELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINVENT AS.
 PA (DZIE-) DZIEGLEWSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAERVIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 XX
 Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX
 WPI; 2001-557614/62.
 DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
 DR AAE10149, AAE10150.
 XX
 PT New nystatin polyketide synthase polynucleotides and polypeptides;

PT useful as antibiotics and antifungals -
 XX
 PS Claim 1; Page 188-254; 266pp; English.
 XX
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrolide antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
 XX
 SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
 Query Match 36.2%; Score 93.4; DB 22; Length 125401;
 Best Local Similarity 63.8%; Pred. No. 3.7e-11;
 Matches 164; Conservative 0; Mismatches 81; Indels 12; Gaps 1;
 QY 1 CTGACCGCGCTGTCCCCACGACGACGAGTCGCTCTCTGGACCTGGTGGCGGCCAC 60
 DB 50584 CTGGCCGCCCTGGACGCGCGCATCAGGCGCTGCTCGACCTGGTCCGACCGAG 50643
 QY 61 ACCATGGCGGTGCTGAACGACGACGAAACAGAGCGACCGGTCGGATGCGGCCCATCG 120
 DB 50644 ACGGCGCGGTGCTCGGCCACTCCGGGCGGACGCCGCTCCGGCGGCGGAC----- 50697
 QY 121 GCGAGTTTCGGCCACCTCGGCTTCGACTTCGATCGGTGTCGAACTGCGCAACCGCCTC 180
 DB 50698 -----TTCAACCGGCTGGCTTCGACTCGGTGATGCGGTGCGAATCGGACGCGGTG 50751
 QY 181 AGCAAGGCCACGCGCTCGGTTGCCCGTAGGCTCATCTTCGACCAACACGCGCGCC 240
 DB 50752 GCCACCGCACCGGAGCCCGCTGCGGCCACGCTCGTCTTCGACCAACCGCGGAC 50811
 QY 241 GCGGTGCGCGCGCGCT 257
 DB 50812 GCGGTGCGCGCGCACCT 50828
 RESULT 6
 AAT80413
 ID AAT80413 standard; DNA; 43280 BP.
 XX
 AC AAT80413;
 XX
 DT 27-FEB-1998 (first entry)
 XX
 DE Ty lactone synthase gene cluster.
 XX
 KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;
 KW polyketide; ty lactone synthesis; antibiotic; tylosin; ss.
 XX
 OS Streptomyces fradiae.
 PH Key Location/Qualifiers
 FT CDS 816..14243
 FT /tag= a
 FT /transl_except= (pos: 816..818, aa: Met)
 FT /note= "ORF1 encodes protein shown in AAW22601"
 FT CDS 14351..19945
 FT /tag= b
 FT /transl_except= (pos: 14351..14353, aa: Met)
 FT /note= "ORF2 encodes protein shown in AAW22602"
 FT CDS 20010..31199
 FT /tag= c
 FT /transl_except= (pos: 20010..20012, aa: Met)
 FT /note= "ORF3 encodes protein shown in AAW22603"
 FT CDS 31232..36067
 FT /tag= d
 FT /note= "ORF4 encodes protein shown in AAW22604"
 FT CDS 36249..41774
 FT /tag= e
 FT /note= "ORF5 encodes protein shown in AAW22605"
 XX
 PN EP791655-A2.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI: 2001-582053/55.
DR P-PSDB; AAG5368.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
XX drugs and pesticides -
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;
Query Match 100.0%; Score 258; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 7e-46;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGACGGCGCTGCCCCACGACGAGCAGTGTCTCTCTGACCTGTGTGGCGCCAC 60
DB 1096 CTGACGGCGCTGCCCCACGACGAGCAGTGTCTCTCTGACCTGTGTGGCGCCAC 1155
QY 61 ACCATGGCGGTCTGAACACGACGAGAACGAGCGCGTCTGGATCGCGGCCATCG 120
DB 1156 ACATGGCGGTCTGAACACGACGAGAACGAGCGCGTCTGGATCGCGGCCATCG 1215
QY 121 GCGAGTTTCGCCCACTCGGCTTCGACTCCGTTCATGGGTGTGAACTGCGCAACCGCTC 180
DB 1216 GCGAGTTTCGCCCACTCGGCTTCGACTCCGTTCATGGGTGTGAACTGCGCAACCGCTC 1275
QY 181 AGCAAGGCCACGGCGTGGGTTCGCCGTGAGCTCATCTTCGACACACACGCGGCC 240
DB 1276 AGCAAGGCCACGGCGTGGGTTCGCCGTGAGCTCATCTTCGACACACACGCGGCC 1335
QY 241 GCGGTCCGCGCGCGCTT 258
DB 1336 GCGGTCCGCGCGCGCTT 1353
RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
XX AAZ58381;
XX
DT 23-MAY-2000 (first entry)
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
DE
KW Polyketide synthase; avermectin; insecticide; ss.
XX Streptomyces avermitilis.
XX
XX WO200001827-A2.
XX
XX 13-JAN-2000.
XX
XX 06-JUL-1999; 99WO-GB02158.
XX
XX 06-JUL-1998; 98GB-0014622.
XX
XX (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX

PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI: 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
XX Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which I
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
Query Match 100.0%; Score 258; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 7e-46;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTGACGGCGTGTCCCCACGACGAGCAGTGTCTCTGACCTGTGTGGCGCCAC 60
DB 1913 CTGACGGCGTGTCCCCACGACGAGCAGTGTCTCTGACCTGTGTGGCGCCAC 1972
QY 61 ACCATGGCGGTCTGAACACGACGAGAACGAGCGCACCGGTCTGGATCGCGGCCATCG 120
DB 1973 ACCATGGCGGTCTGAACACGACGAGAACGAGCGCACCGGTCTGGATCGCGGCCATCG 2032
QY 121 GCGAGTTTCGCCCACTCGGCTTCGACTCCGTTCATGGGTGTGAACTGCGCAACCGCTC 180
DB 2033 GCGAGTTTCGCCCACTCGGCTTCGACTCCGTTCATGGGTGTGAACTGCGCAACCGCTC 2092
QY 181 AGCAAGGCCACGGCGTGGGTTCGCCGTGAGCTCATCTTCGACACACACGCGGCC 240
DB 2093 AGCAAGGCCACGGCGTGGGTTCGCCGTGAGCTCATCTTCGACACACACGCGGCC 2152
QY 241 GCGGTCCGCGCGCGCTT 258
DB 2153 GCGGTCCGCGCGCGCTT 2170
RESULT 3
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
XX AAA92301;
XX
DT 10-JAN-2001 (first entry)
DE
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
XX agrochemical; ds.
XX Streptomyces avermitilis.
XX
XX Key Location/Qualifiers
FH 1..11919
FT CDS /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 51.3231 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353

Perfect score: 258

Sequence: 1 ctgacgcgctgtcccccac.....ccgcggtcgccgcgcctt 258

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125995159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	258	100.0	11916	22	AAH79279
2	258	100.0	12381	21	AAZ58381
3	258	100.0	30690	21	AAH92301
4	258	100.0	30690	22	AAH79277
5	93.4	36.2	125401	22	AAH7186
6	91.8	35.6	43280	18	AAH80413
7	90.2	35.0	53789	19	AAV21187
8	85.4	33.1	5676	19	AAV21186
9	83.8	32.5	20394	22	AAF24892

C	10	83.2	32.2	77536	21	AAA14651	Nucleotide sequenc
	11	82.6	32.0	2700	22	AAH7193	ERD48 insert DNA i
	12	82.6	32.0	65140	22	AAH7184	Streptomyces nous
	13	82.4	31.9	5544	22	AAF24891	Pimaricin biosynth
	14	82.2	31.9	44377	18	AAH78508	Platenolide synth
	15	82.2	31.9	44377	18	AAH78508	Platenolide synth
	16	79.4	30.8	11220	21	AAZ87298	S. venezuelae macr
	17	79.4	30.8	36778	21	AAZ87318	S. venezuelae pik
	18	79.4	30.8	37948	21	AAZ87285	S. venezuelae pik
	19	79.4	30.8	38506	21	AAH75633	Nucleotide sequenc
	20	79.4	30.8	38506	21	AAZ58001	Recombinant cosmid
	21	77.4	30.0	1681	24	AAH18438	Contig 115 DNA enc
	22	77.4	30.0	29879	14	AAQ46806	eryA region of S.
	23	77.4	30.0	31422	22	AAH92302	S. avermitilis ave
	24	77.4	30.0	31422	22	AAH79278	Streptomyces averm
	25	74.6	28.9	990	21	AAH55784	Type I polyketide
	26	74.6	28.9	12449	21	AAH55840	Complete Mitomycin
	27	74.6	28.9	13842	21	AAZ87297	S. venezuelae macr
	28	74.6	28.9	18331	21	AAH55857	Complete nucleotid
	29	73.6	28.5	4851	22	AAH52062	Mycobacterium tube
	30	73.6	28.5	4403765	22	AAH99683	Mycobacterium tube
	31	73.6	28.5	4411529	22	AAH99682	Mycobacterium tube
	32	73.4	28.4	16767	22	AAH83339	S. spinosa DNA fra
	33	73.4	28.4	25360	22	AAH83317	S. spinosa DNA fra
	34	73.4	28.4	29736	22	AAH8317	S. spinosa DNA fra
	35	73.4	28.4	80161	20	AAZ21501	DNA fragment of Sa
	36	73.2	28.4	15872	21	AAZ87283	S. venezuelae vep
	37	72.8	28.2	748	22	AAH89129	Polyketide synthas
	38	72.8	28.2	753	22	AAH89143	Polyketide synthas
	39	72.6	28.1	1434	24	AAH18434	Contig 132 DNA enc
	40	71.6	27.8	47981	22	AAH30757	Micromonospora meg
	41	71.4	27.7	27541	22	AAH17185	Streptomyces nous
	42	71.2	27.6	15872	18	AAH68715	Streptomyces venez
	43	71	27.5	77536	21	AAA14651	Nucleotide sequenc
	44	70.4	27.3	561	22	AAH89135	Polyketide synthas
	45	70.2	27.2	1871	24	ABK11624	S. erythraea monen

ALIGNMENTS

RESULT 1

AAH79279
ID AAH79279 standard; DNA; 11916 BP.

XX AAH79279;

XX 04-DEC-2001 (first entry)

DT Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.

DE Avermectin aglycone synthase; AAS; avermectin derivative;

XX drug production; veterinary drug; pesticide; ds.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 1..11916

FT /*tag= a

FT /product= "AAG65268"

FT /partial

XX WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP01381.

XX 24-FEB-2000; 2000JP-0047405.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX (KITA) KITASATO INST.

```
/ APPLICANT: Broughton, Mary C
/ APPLICANT: Crawford, Kathryn P
/ APPLICANT: Madduri, Krishnamurthy
/ APPLICANT: Treadway, Patti J
/ APPLICANT: Turner, Jan R
/ APPLICANT: Waldron, Clive
/ TITLE OF INVENTION: Bio-synthetic Genes For Spinosyn Insecticide
/ FILE REFERENCE: 50489 DIV1
/ CURRENT APPLICATION NUMBER: US/09/370,700
/ CURRENT FILING DATE: 1999-08-09
/ EARLIER APPLICATION NUMBER: US 09/36987
/ EARLIER FILING DATE: 1998-03-09
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 80161
/ TYPE: DNA
/ ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

Query Match      28.4%; Score 73.4; DB 4; Length 80161;
Best Local Similarity 65.6%; Pred. No. 1.5e-08;
Matches 107; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 95 GCACCGCGTCGGATCGCGCCCATCGCGAGTTTCGCCACCTCGGCTTCGACTCCGTCA 154
Db 75497 GCAGGTCGAGCGATGCGACATCGAGAAGCTTTGGCCGAGCTGGGTTTCGACTCGCTGA 75556

QY 155 TGGGTGTCGAACTCGCCAACCCGCTCAGCAAGCCACGGGCTCGGTTGCCCGTGACGC 214
Db 75557 CGGCCATCGAACTCCGCCAACCCGCTCGTACCGCCACCGGACTCGGCTTCCCGGACGC 75616

QY 215 TCATCTTCGACACACACCGCGCGCGTGC CGCGCGCT 257
Db 75617 TGGCCTTCACACACGCACTGCGCGCGCACTGCCCCAGCACGT 75659

RESULT 15
US-09-105-537-1
/ Sequence 1, Application US/09105537A
/ Patent No. 6265202
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/105,537A
/ CURRENT FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1
/ LENGTH: 15872
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match      28.4%; Score 73.2; DB 4; Length 15872;
Best Local Similarity 64.1%; Pred. No. 1.5e-08;
Matches 127; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY 63 CATGGCGGTGCTGAACGACGACGAAACGAGCGCACCGGTGCGATGCGGGCCCATCGGC 122
Db 8203 CGTGACGGGTGTCGTGCGGAGCTCTCGGGACGCGGACTCGCCGCGGTGGGGCGGA 8262

QY 123 GAG---TTTCGCCCACTCGGTTTCGACTCCGTCATGGGTGTCGAATCGCAACCGCCT 179
Db 8263 GCGGTCTTCAAGGACCTCGGCTTCGACTCCCTGGCGGGGTGGAGCTCCGCAACCGGCT 8322

QY 180 CAGCAGGCGCACGGGCTCGGTTGCCGTGACGCTCATCTTCGACCACACACCGCGGC 239
Db 8323 GAACGCGGCACCGGCTCGGCTCCCGCGACACCGGTCTTCGACCATCCCTCGCGGC 8382
```

```
QY 240 CGCGGTGCGCGCGCCT 257
Db 8383 CGCGGTGCGGTCCCATCT 8400
```

Search completed: June 18, 2003, 01:06:20
Job time : 29.1763 secs

APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 28.5%; Score 73.6; DB 4; Length 4403765;
Best Local Similarity 66.2%; Pred. No. 1.7e-08;
Matches 106; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 93 GGCACCGGTGCGATCGCGGCCCATCGCGAGTTTCGCCACCTCGGCTTCGACTCCGT 152
Db 3286037 GCCTCCGCGAGGAGCTCGACCCGACACCGAAATTCGCGACCTCGGTTTCGACTCAT 3285978

QY 153 CATGGGTGTCGAACCTCGCAACCGCTCAGCAAGGCCACGGGCTCGGTTGCCCGTGAC 212
Db 3285977 AACGGCTGTGGATTAGCAACCGCTCAAAACCGCACCGGACTGACGCTGCCACTAC 3285918

QY 213 GCTCATCTTCGACCAACACCGCCGCGCGCGCGGTCGCGCG 252
Db 3285917 CGTGATTTTCGATCATCCCACTCCCACTCGCGTTCGCGCGAG 3285978

RESULT 12
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 28.5%; Score 73.6; DB 4; Length 4411529;
Best Local Similarity 66.2%; Pred. No. 1.7e-08;
Matches 106; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 93 GGCACCGGTGCGATCGCGGCCCATCGCGAGTTTCGCCACCTCGGCTTCGACTCCGT 152
Db 3291761 GCCTCCGCGAGGAGCTCGACCCGACACCGAAATTCGCGGACCTCGGTTTCGACTCAT 3291702

QY 153 CATGGGTGTCGAACCTCGCAACCGCTCAGCAAGGCCACGGGCTCGGTTGCCCGTGAC 212
Db 3291701 AACGGCTGTGGATTAGCAACCGCTCAAAACCGCACCGGACTGACGCTGCCACTAC 3291642

QY 213 GCTCATCTTCGACCAACACCGCCGCGCGCGGTCGCGCG 252
Db 3291641 CGTGATTTTCGATCATCCCACTCCCACTCGCGTTCGCGCGAG 3291602

RESULT 13

US-09-036-987A-1
Sequence 1, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
NUMBER OF SEQ ID NOS: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80161 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-036-987A-1

Query Match 28.4%; Score 73.4; DB 3; Length 80161;
Best Local Similarity 65.6%; Pred. No. 1.5e-08;
Matches 107; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 95 GCACCGCTCGGATCGCGGCCCATCGCGGAGTTTCGCCACCTCGGCTTCGACTCCGTCA 154
Db 75497 GCAGTCTGAGCGATGTGCACATCGAGAGCCCTTCGCCGAGCTGGGTTTCGACTCGCTGA 75556

QY 155 TGGGTGTGAACTGCGCAACCGCTCTCAGCAAGGCCACGGGCTCGGTTGCCCGTGACGC 214
Db 75557 CGGCCATCGAACTCGCGCAACCGCTCTCGCTACCGCCACCGGAGTGGGGTTCCCGGACGC 75616

QY 215 TCATCTTCGACCAACACCGCGCGCGGCTCGCGCGCGCT 257
Db 75617 TGGCTTCGACCAACCGCGGCTCGCGCGCGCT 75659

RESULT 14
US-09-370-700-1
Sequence 1, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H


```

; LOCATION: 744..6659
; OTHER INFORMATION: /functions= "APPROXIMATE SPAN OF
; OTHER INFORMATION: MODULE 1"
; OTHER INFORMATION: /label= FUNCTION
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 744..11219
; OTHER INFORMATION: /functions= "gene= "eryA""
; OTHER INFORMATION: /products= "ORF1 encoding modules 1 & 2 for
; OTHER INFORMATION: 6-deoxyerythronolide B"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 744..1868
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyltransferase domain 1 of module 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1998..2198
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyl carrier domain 1 of module 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2250..3626
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3831..4811
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyltransferase 2 domain of module 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5574..6125
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6369..6626
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyl carrier domain 2 of module 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6678..11219
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: module 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 6678..8066
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 8262..9305
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyltransferase domain of module 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 9506..10454
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 10707..10964
; OTHER INFORMATION: /functions= "approximate span of
; OTHER INFORMATION: acyl carrier domain of module 2"
;
US-08-439-009A-1

```

```

Query Match 30.0%; Score 77.4; DB 3; Length 11219;
Best Local Similarity 59.9%; Pred. No. 1.6e-09;
Matches 154; Conservative 0; Mismatches 91; Indels 12; Gaps 1;

QY 1 CTGACCGCGCTGCCCCACGAGCAGTGCTCTCTGACACTGTGGCGGCCAC 60
|||||

```

```

Db 10707 CTGGCGGCTCTGCGCGTCCGAGCGGAGCTGGTCCGCTGGTCCGACCCAGC 10766
QY 61 ACCATGGCGGTGCTGAACGACGACGGAACGAGCGCACCGGTCGGATGCGGCCCATCG 120
|||||
Db 10767 ACCGCGACCGTCTGGCCACGACG-----CCGAAAGCGGTGCGCGGACC 10814
QY 121 GCGAGTTTTCGCCACCTCGGCTTCGACTCCGTCATGGGTGTGAACTCGGCAACGCTTC 180
|||||
Db 10815 ACGCGTTTCAAGGAGCTCGGTTTCGACTCGCTGGGCGCGCTCGGCTGCGCAACCTGCTC 10874
QY 181 AGCAAGGCGACGGGCTCGGTTGCGGCTGAGCTCATCTTCGACACACACGCGGCC 240
|||||
Db 10875 AACGCGGCACCGGGCTCGGCTCGGCTCGAGCTGGTCTTTCGACACCCGAAACGCTCC 10934
QY 241 GCGGTGCGCGCGCGCT 257
|||||
Db 10935 GCGGTGCGCGGTTTCCT 10951
|||||

```

RESULT 10

```

US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, Y.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

```

```

Query Match 28.9%; Score 74.6; DB 4; Length 13842;
Best Local Similarity 55.6%; Pred. No. 7.1e-09;
Matches 143; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

```

```

QY 1 CTGACCGCGCTGTCCCCACGACGAGCTCGCTCCGCGAGCTGCTCTCTCGTCTC 60
|||||
Db 7441 CTGGCCGAGCGCTGGCGCGCGCGCTCCGCGAGCTGCTCTCTCGTCTC 7500
QY 61 ACCATGGCGGTGCTGAACGACGACGGAACGAGCGCACCGGTCGGATGCGGCCCATCG 120
|||||
Db 7501 GTACGGGCGCAGCGCGCGCGCTCCGATGCTCCGCGAGGAGCTCGCGCGGAC 7560
QY 121 GCGAGTTTTCGCCACCTCGGTTTCGACTCCGTCATGGGTGTGAACTGCGCAACGCTTC 180
|||||
Db 7561 CGCGCTTCAAGGACATCGGTTTCGACTCGCTCGCGGTGTGAGCTGCGCAACAGGCTG 7620
QY 181 AGCAAGGCGACGGGCTCGGTTGCGGCTGAGCTCATCTTCGACACACACGCGGCC 240
|||||
Db 7621 ACCCGGGGACCGGGCTCAGCTCCCGGACGCTCTTCGACACCCGAGCGCGCTG 7680
QY 241 GCGGTGCGCGCGCGCT 257
|||||
Db 7681 GCGCTCGTCTGCTGCT 7697
|||||

```

RESULT 11

```

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

```

OTHER INFO
; US-07-642-734C-1

```

; STRAIN: NRRL 2338
;
; FEATURE:

```

Patent No. 6117659
GENERAL INFORMATION:
APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 38506
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 30.8%; Score 79.4; DB 3; Length 38506;
Best Local Similarity 60.2%; Pred. No. 6.2e-10;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCACGACGACATGCTGCTCTGTGACCTGTGTGCGCGCCACACCATGGC 68
DB 24506 GCTCGCGCGATGACACCGGACGACCGGGTTCGCGACCTGCGGGACCTCGTACGCA 24565
QY 69 GTGTGTAACGACGAGGAAAGCGGACCGGTCGGATGCGGCCCATCGCGAGTTT 128
DB 24566 CTTGGCGACCGTCTCTGGGACGCG---GCACCCGAGCGCGGTGGACCTGGAGCGGCTT 24622
QY 129 CGCCCACTCGGCTTCGACTCGCTCATGGGTGTGCAACTGCGCAACCGCTCAGCAAGC 188
DB 24623 CGCGACACCGGTTTCGACTCGCTCATCGCGCTCGACTCGCAACCGCTCTCAACGCGC 24682
QY 189 CACGGCCTCGGTTGCCGTGACGCTCATCTTCGACCAACACACCGCGCGCGGTGCG 248
DB 24683 GACCGGCTCGGCTGCGCGCACGCTGCTTTCGACCAACACCGCGCGGAGTCTCG 24742
QY 249 CGCGCGCCT 257
DB 24743 CGGGCACCT 24751

RESULT 7
US-09-434-288-7
Sequence 7, Application US/09434288
Patent No. 6303767
GENERAL INFORMATION:
APPLICANT: Betlach C., Melanie
APPLICANT: MCDANIEL, Robert
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
FILE REFERENCE: 30062-20030.00
CURRENT APPLICATION NUMBER: US/09/434,288
CURRENT FILING DATE: 1999-11-05
PRIOR APPLICATION NUMBER: 60/107,093
PRIOR FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7

LENGTH: 1681
TYPE: DNA
ORGANISM: Streptomyces narbonensis
US-09-434-288-7

Query Match 30.0%; Score 77.4; DB 4; Length 1681;
Best Local Similarity 59.9%; Pred. No. 1.5e-09;
Matches 154; Conservative 0; Mismatches 91; Indels 12; Gaps 1;

QY 1 CTGACCGCGCTGTCCTCCACGACGACGACGCTGCTCTGACCTGGTGGTGGCGGCCAC 60
DB 541 CTCGCGCGATGACCGCGGACGACGACGACGACGACGACGACGACGACGACGACG 600
QY 61 ACCATGGCGGTGCTGAACGACGACGACGACGACGACGACGACGACGACGACGACG 120
DB 601 GTGGCGACCGTCTGGGACACG-----GGGCCCCGAGCGCGGTGACCTGGAG 648
QY 121 CGAGTTTCGCCACCTCGGCTTCGACTCCCTCATGGGTGTCGAACTCGGCAACGCGCTC 180
DB 649 CGCGCTTTCGCGACACCGGTTTCGACTCCCTCACCGCGCTCGAGCTCCGCAACGCGCTC 708
QY 181 AGCAAGCGCACGGCGCTGCGGTTCGCGTACGCTCATCTTCGACCAACACGCGCGCC 240
DB 709 AACCGCGCACCGGCTGCGCTTCGCGCACGCTCTTCGACCAACCACTCCGGGG 768
QY 241 CGCGTTCGCGCGCGCT 257
DB 769 GAGCTCGCGCGCACCT 785

RESULT 8
US-07-642-734C-1
Sequence 1, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 35.6%; Score 91.8; DB 2; Length 43280;
Best Local Similarity 63.4%; Pred. No. 9.8e-13;
Matches 163; Conservative 0; Mismatches 82; Indels 12; Gaps 1;
QY 1 CTGACCGCGCTGCCCCAGCAGCAGTCTCTCTGAGACCTGTGTGGCGGCCAC 60
Db 30648 CTGACCGCGCTCAGCGCAGAGGAGCGCCCTCTCTGCGCCAGATCGCGGCCAC 30707
QY 61 ACATGCGCGTCTGAACGACGAGGAAACGAGCGCACCGCTCGGATCGCGGCCATCG 120
Db 30708 GCGCGCGCGTCTGGGCGACCGCTCCGACGACTCGCCGAGGACCGGGCC----- 30761
QY 121 GCGAGTTTCGCCACCTCGCTTCGACTCCGTCATCGGTGTCGAACCTGGCGCAACCGCTC 180
Db 30762 -----TTCAAGGACCTCGCTTCGACTCGCTCACCGCGTGGAGATGGCAACCGCTG 30815
QY 181 AGCAAGCGCAGCGCGCTGCGGTGCGGTGAGCGCTCATTTGACACACACCGCGCGCC 240
Db 30816 AGCGCGCGCACCGCGCTCGCGTGGCGCCACCGCTCTCTGACCAACCGACCGCGCGC 30875
QY 241 GCGGTGCGCGCGCGCT 257
Db 30876 GAGTGGCGCGCACCT 30892

RESULT 2
US-08-804-227C-7
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7
Query Match 31.9%; Score 82.2; DB 2; Length 44377;
Best Local Similarity 61.1%; Pred. No. 1.5e-10;
Matches 157; Conservative 0; Mismatches 88; Indels 12; Gaps 1;
QY 1 CTGACCGCGCTGTCCCCACGACGAGCAGTCGCTGCTCTGAGACCTGTGTGGCGGCCAC 60
Db 19476 CTCGCGGAGAGTTCCGACCCGAGGAGTACGGCTCTCTGCTCGACTGTGTGGGACGAG 19535
QY 61 ACCATGGCGGTCTGAACGACGACGAAACGAGCGCACCGGTCTCGGATGCGCGGCCATCG 120
Db 19536 GTGCGCGCGGTCTCGGACAGC-----GCGACCCCGCGCATCGCGCGCGC 19583
QY 121 GCGAGTTTCGCCACCTCGGTTTCGACTCCGTCATGGGTGTCGAACCTGGCGCAACCGCTC 180
Db 19584 GCGACCTTCAAGGACGCGCGGATTGACCTCCCTCACCGTGTGACCTCGCGAACCGGCTG 19643
QY 181 AGCAAGGCGCACCGCGCTGCGGTTCGCGTTCGAGCTCATCTTCGACCAACACACGCGCGC 240
Db 19644 AACACACGACCGGACTGGGCTGCGCGGACCTCTCTGACCAACACACGCGCTC 19703
QY 241 GCGGTGCGCGCGCGCT 257
Db 19704 GCGCTGCGCGAAGTCT 19720
RESULT 3
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 11.1763 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353

Perfect score: 258

Sequence: 1 ctgacgcgctgtcccccac;.....ccgcggtgcgcgcgcctt 258

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338391 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/5A COMB.seq:*

2: /cgn2_6/prodata/1/ina/5B COMB.seq:*

3: /cgn2_6/prodata/1/ina/6A COMB.seq:*

4: /cgn2_6/prodata/1/ina/6B COMB.seq:*

5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq:*

6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.8	35.6	43280	2	US-08-804-227C-1
2	82.2	31.9	44377	2	US-08-804-227C-7
3	82.2	31.9	44377	2	US-08-804-198-1
4	79.4	30.8	11220	4	US-09-105-537-32
5	79.4	30.8	36778	4	US-09-105-537-5
6	79.4	30.8	38506	3	US-09-320-878-19
7	77.4	30.0	1881	1	US-09-434-288-7
8	77.4	30.0	11219	1	US-07-642-734C-1
9	77.4	30.0	11219	3	US-08-439-009A-1
10	74.6	28.9	13842	4	US-09-105-537-30
11	73.6	28.5	4403765	4	US-09-103-840A-2
12	73.6	28.5	4411529	4	US-09-103-840A-1
13	73.4	28.4	80161	3	US-09-036-987A-1
14	73.4	28.4	80161	4	US-09-370-700-1
15	73.2	28.4	15872	4	US-09-105-537-1
16	72.8	28.2	748	3	US-09-154-083-1
17	72.8	28.2	753	3	US-09-154-083-28
18	72.6	28.1	1434	4	US-09-434-288-3
19	70.6	27.4	20235	1	US-07-642-734C-3
20	70.6	27.4	20235	3	US-08-439-009A-3
21	70.4	27.3	561	3	US-09-154-083-13
22	69	26.7	50937	4	US-09-428-517-1
23	68.6	26.6	4403765	4	US-09-103-840A-2
24	68.6	26.6	4411529	4	US-09-103-840A-1
25	68.4	26.5	4689	4	US-09-105-537-34
26	67.6	26.2	13987	2	US-08-804-227C-13
27	64.6	25.0	33529	4	US-09-144-085-3

28	63.2	24.5	28958	1	US-08-258-261B-6	Sequence 6, Appli
29	63.2	24.5	28958	1	US-08-456-837-6	Sequence 6, Appli
30	63.2	24.5	28958	1	US-08-457-342-6	Sequence 6, Appli
31	63.2	24.5	28958	1	US-08-457-646A-6	Sequence 6, Appli
32	63.2	24.5	28958	1	US-08-458-076A-6	Sequence 6, Appli
33	63.2	24.5	28958	1	US-08-764-233A-4	Sequence 4, Appli
34	63.2	24.5	28958	1	US-08-457-335A-6	Sequence 6, Appli
35	63.2	24.5	28958	1	US-08-729-214-6	Sequence 6, Appli
36	63.2	24.5	28958	3	US-09-028-934-6	Sequence 6, Appli
37	63.2	24.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
38	58.6	22.7	4041	4	US-09-105-537-36	Sequence 36, Appli
39	-55	21.3	68750	3	US-09-335-409-1	Sequence 1, Appli
40	55	21.3	68750	4	US-09-568-102-1	Sequence 1, Appli
41	55	21.3	68750	4	US-09-567-969-1	Sequence 1, Appli
42	55	21.3	68750	4	US-09-568-480-1	Sequence 1, Appli
43	55	21.3	68750	4	US-09-568-486-1	Sequence 1, Appli
44	55	21.3	68750	4	US-09-568-472-1	Sequence 1, Appli
45	55	21.3	68750	4	US-09-567-899-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:

Db 7561 CCGCCTTCAAGACATCGCTTCACTGCTGCGCGGTGAGTGGCAACGCGCTG 7620
Qy 181 AGCAAGGCAAGGCGCTGCGGTGCGCTGAGCGCTCATTTGACCAACGCGCGCTC 240
Db 7621 ACCCGGCGACCGGCGCTGCGCGCGCGAGCGCTGCTTGCACCAACGCGCGCTG 7680
Qy 241 GCGGTGCGCGCGCGCT 257
Db 7681 GCGCTGCTGCTGCTGCT 7697

RESULT 14
US-09-712-363-116
Sequence 116, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
APPLICANT: Eisenberg, David
APPLICANT: Rotstein, Sergio H.
APPLICANT: Marcotte, Edward M.
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
FILE REFERENCE: 07419-032001
CURRENT APPLICATION NUMBER: US/09/712,363
PRIOR FILING DATE: 2000-11-13
PRIOR APPLICATION NUMBER: PCT/US00/02246
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: 60/179,531
PRIOR FILING DATE: 2000-02-01
PRIOR APPLICATION NUMBER: 60/117,844
PRIOR FILING DATE: 1999-01-29
PRIOR APPLICATION NUMBER: 60/118,206,
PRIOR FILING DATE: 1999-02-01
PRIOR APPLICATION NUMBER: 60/126,593
PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 116
LENGTH: 4851
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-116

Query Match 28.4%; Score 73.6; DB 9; Length 4851;
Best Local Similarity 66.2%; Pred. No. 6,8e-12;
Matches 106; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 93 GCGCAGCGGCTGCG 152
Db 4590 GCGCTCGCGCGAGAGCGTCAACCCGACACCGAATTCGGGACCTCGGTTGACATCATT 4649
Qy 153 CATGGGTGTCGAACCTGCGCAACCGCTCAAGCAAGCCAGCGGCGTGGCGCGTAC 212
Db 4650 AACGGCTGTGAGATTAGCAACCGCTCAAAACCGCACCGAGACTGACGCTGACACTAC 4709
Qy 213 GCTCATCTTGAACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
Db 4710 CGTGATTTTCATCATCTCCACTCCACTGCGGTGCGCGAG 4749

RESULT 15
US-09-860-846-1
Sequence 1, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:

APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIOR FILING DATE: 2001-05-18
CURRENT APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 28.4%; Score 73.2; DB 9; Length 15872;
Best Local Similarity 64.1%; Pred. No. 8,6e-12;
Matches 127; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

Qy 63 CATGCGGTGCTGAACGACGACGGAAGCGCACCGCGTGGATCGCGCGCGCATCGGC 122
Db 8203 CGTGAAGGATGTCGCGGAGCGCTCGGCGACCGCACTCCGCGCGGTGCGGCGGA 8262
Qy 123 GAG--TTTCCGCCACTCGCGCTTGAATCCGTATGCTGTCGCACTGCGCAACGCGCT 179
Db 8263 GCGGTCTTCAAGACCTCGCTTGAATCCGTGCGCGCGGTGCGGAGCTCGCAACGCGCT 8322
Qy 180 CAGCAAGCCAGCGGCGCTGCGGTTGCCGTGACGCTCATCTTGCACACACCGCGCGC 239
Db 8323 GAACCGCGCACCGGCGCTCGCGCTCCCGCACACAGGCTTTCGACATCCCTCGCGCGC 8382
Qy 240 GCGGTGCGCGCGCGCT 257
Db 8383 CCGCTCGCGTCCCATCT 8400

Search completed: June 18, 2003, 05:08:05
Job time : 31.5809 secs

OY 121 GCGAGTTTCGCCACCTCGGGCTTGACATCCGTCATGGGTGTGAACTGCGGAAACCGCTC 180
 Db 7561 GCGCGCTTCAAAGAAATCGGCTTCACATCGCTCGCCGGTGTGTGACATCGGGAACAAGGCTG 7620
 OY 181 AGCAAGGCGACCGGGCTCGGTGCGCGTGAAGCTCATCTTCGACACCAACGACGCGGCGC 240
 Db 7621 ACCCGGGGACCGGGCTCCAGCTGCGCCCGAAGCTGTCTTGACCAACCGACGCGCGCTG 7680
 OY 241 GCGGTGCGCGCGCGGCT 257
 Db 7681 GCCCTCGTGTGCTGCT 7697

```

RESULT 11
US-09-968-384B-30
Sequence 30, Application US/09988384B
Publication NO. US20030073824A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and plikromycin
FILE REFERENCE: 600,536US1
CURRENT APPLICATION NUMBER: US/09/968,384B
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: PCT/US99/14398
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-968-384B-30

```

Query Match	Best Local Similarity	28.9%;	Score 74.6;	DB 9;	Length 13842;
Matches 143;	Conservative	0;	Mismatches 114;	Indels	Gaps 0;
Qy	1	CTGACCGCGCTGTGTCCTCCCGACGACGACGATGCGTCTCTGGAACCTGGTGGCGCCAC	60		
Db	7441	CTGGCCGACGCGCGCTGGCGCGCGCGCTCCCGACGAGCTACGAGGATCTCTCTGGTCTC	75000		
Qy	61	ACCATGGCGGAGTCTGAACGACGACGGAACGAGCGACCCGCGTGGAGATGCGCGGCATCG	120		
Db	7501	GTACGGGGGACGAGCGCGCGCGCTGCTCCGAGTGGCGTTGCGCGAGAGAGTGGCGCGAC	7560		
Qy	121	GCGAGTTTCGCCCACTGGCGCTTGACTCCGTCAATGGGTGTGAACTGGCGCAACGCGCTC	180		
Db	7561	CGCGCTTTCAGAGACATCGGCTTGACTCGCTGCCGAGTGCAGCTGCGCGAACAGCTG	7620		
Qy	181	AGCAAGGSCACGAGGCGCTGCGGTTCCCGGTGACGCTCACTTCAGACCAACACAGCGGACC	240		
Db	7621	ACCGGGGAGACGGGGCTTCAGCTCCCGCGACGCTCGCTTCGACACACCGACCGCCCTG	7680		
Qy	241	GCGGTGCGCGCGCGCT	257		
Db	7681	GCCCTCGATGCTGCT	7697		

RESULT 12
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US2003008705A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U01

```

; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

```

Query Match	28.9%	Score 74.6	DB 9	Length 13842
Best Local Similarity	55.6%	Pred. No. 3.5e-12		
Matches 143	Conservative 0	Mismatches 114	Indels 0	Gaps 0
Qy	1	CTGACCCCGCGCTGTCTCCCCACGCACACAGTCGCTGCTCTTGACCTGTGCGGCCAC	60	
Db	7441	CTGGCCCAAGCGGCTGGGCGCGCGGCGCTCCCGGCGAGGTACGAGATCTCTTGCTCTC	7500	
Qy	61	ACCATGGCGGTGTCTGAACGACGACGAAACGAGCGCACCGCGTGGATGCGCGCCATCG	120	
Db	7501	GTACGGGCGGACGGCGCGCGCGTCTCCGGATGCTGTGGCGAGAGACGTGCGCCCGAC	7560	
Qy	121	GCGAGTTTCGCCCACTTCGCGCTTCGACTCCGTCAATGAGTGTGCACTGCGCAACCGCTCTC	180	
Db	7561	CGCCCTTTCAAAGACATCGGCTTCGACTCGCTCGCGGTGTCAGCTGCGCAACGAGCTG	7620	
Qy	181	AGCAAGGCCACGCGGCTTCGCGTTCGCCGTGACGTGATCTTTCCACCAACACACGCGCGGCC	240	
Db	7621	ACCGGGCGACCGGGCTTCAGCTGCGCCGCGACGCTCGTCTTCGACCAACCGACCGCGCTG	7680	
Qy	241	GCGGTCGCGCGCGCGCT	257	
Db	7681	GCCCTCGTGTGCTGCT	7697	

```

RESULT 13
US-09-861-289-30
; Sequence .30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-30

```

[illegible]

Db 26601 CGGGCAGCT 26609

RESULT 8

US-09-988-384B-5
Sequence 5, Application US/09988384B
Publication No. US20030073824A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,536US1
CURRENT APPLICATION NUMBER: US/09/988,384B
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: PCT/US99/14398
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 5
LENGTH: 37948
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match 30.8%; Score 79.4; DB 9; Length 37948;
Best Local Similarity 60.2%; Pred. No. 1,5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCCAGCAGAGATGCTGCTCTGAGCTTGTGCGGCGCCACCATGCG 68
DB 27534 GCTGCGCGCATACACCGAGCAGCGGCTGCGCACCTGCGTCTGACGA 27593
QY 69 GGTCTGAACGACGACGAGAAAGAGCGACCGGCTGGATGCGGCCATCGCGATT 128
DB 27594 CTTGGGACCGCTCTGAGACG---GCACCCGACCGGGTGGACTGGAGGGGCTT 27650
QY 129 CGCCCACTCGGCTTGCATCTCGTATGCTGTAATCGCGAACCGCTCGCAAGC 188
DB 27651 CGCGACACCGGTTGCTGACTGCTGACCGCGCTCGAATCGCAACCGTCTCAACCGCG 27710
QY 189 CAGGCGCTGCGGTTGCGGCGCTGATGCTGATGCTGACCAACACCGCGCGCGGTCG 248
DB 27711 GACCGGCTGCGGCTGCGGCGACGCTGCTTTCGACCAACCGCGGCGGAGCTCGC 27770
QY 249 CGCGCGCT 257
DB 27771 CGGGCAGCT 27779

RESULT 9

US-09-793-708-19
Sequence 19, Application US/09793708
Publication No. US20030104597A1
GENERAL INFORMATION:
APPLICANT: Ashley, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002121
CURRENT APPLICATION NUMBER: US/09/793,708
PRIOR FILING DATE: 2001-02-22
PRIOR APPLICATION NUMBER: US 09/657,440
PRIOR FILING DATE: 2000-09-07
PRIOR APPLICATION NUMBER: US 09/320,878
PRIOR FILING DATE: 1999-05-27
PRIOR APPLICATION NUMBER: US 09/141,908
PRIOR FILING DATE: 1998-08-28
PRIOR APPLICATION NUMBER: US 09/073,538

PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/846,247
PRIOR FILING DATE: 1997-04-30
PRIOR APPLICATION NUMBER: US 60/134,990
PRIOR FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 38506
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 30.8%; Score 79.4; DB 9; Length 38506;
Best Local Similarity 60.2%; Pred. No. 1,5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCCAGCAGAGATGCTGCTCTGAGCTTGTGCGGCGCCACCATGCG 68
DB 24506 GCTGCGCGCATACACCGAGCAGCGGCTGCGCACCTGCGGAGCTGTCGTAACGA 24565
QY 69 GGTCTGAACGACGACGAGAAAGAGCGACCGGCTGGATGCGGCCATCGCGATT 128
DB 24566 CTTGGGACCGCTCTGAGACG---GCACCCGACCGGGTGGACTTGAACCGGCGCTT 24622
QY 129 CGCCCACTCGGCTTGCATCTCGTATGCTGTAATCGCGAACCGCTCGCAAGC 188
DB 24623 CGCGACACCGGTTGCTGACTGCTGACCGCGCTGCAACCGTCTCAACCGCG 24682
QY 189 CAGGCGCTGCGGTTGCGGCGCTGATGCTGATGCTGACCAACACCGCGCGGTCG 248
DB 24683 GACCGGCTGCGGCTGCGGCGACCGCTGCTTTCGACCAACCGCGGCGGAGCTCGC 24742
QY 249 CGCGCGCT 257
DB 24743 CGGGCAGCT 24751

RESULT 10

US-09-860-846-30
Sequence 30, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match 28.9%; Score 74.6; DB 9; Length 13842;
Best Local Similarity 55.6%; Pred. No. 3,5e-12;
Matches 143; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1 CTGACCGCGCTGCTCCCGCAGCAGAGATGCTGCTCTGAGCTTGTGCGGCGCAC 60
DB 7441 CTGGCGAGCGGCTGCGCGCGCGCGCTCCCGCGAGCGTAAGAGATCTCTCGGTTC 7500
QY 61 ACCATGCGGCTGCTGAACGACGAGAAAGAGCGACCGGCTGGATGCGGCGCATG 120
DB 7501 GTACGGCGCAGACGCGCGCGCTGCTCGGATGCTTCCCGAGAGACTCGCGCGAC 7560

QY 249 CGCGCGCCT 257
DB 10914 CGGGCACCT 10922

RESULT 5

US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 30.8%; Score 79.4; DB 9; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCCAGCAGCAGCAGTGTGCTGCTCTGAGCTGTGCGGCCACACCATGAGC 68
DB 26364 GCTGCCCGGAGTGAACACCGAGACCGGGTCCGACCTCGTCCGTACGCA 26423
QY 69 GGTGCTGAACGACGAGAAACGAGCGCACCGGTGAGTCCGGCCATCGCGAGATT 128
DB 26424 CGTGGCAGACCGTCTCGGACACG---GCAACCCGAGCCGGGTGACCTGAGAGCGGCTT 26480
QY 129 CGCCCACTCGGCTTGTGACTCCCTCATGGGTGTGCAACTGCGAACCCTCAGCAAGC 188
DB 26481 CGCGACACCGGTTTGACTCGTCAACCGCGTGAATCCGGAACCGTCTCAACCGCCG 26540
QY 189 CACGGGCTGCGGTTGCGCGTCACTCTTGAACACACACCGCGCGCGGTCCG 248
DB 26541 GACCGGCTGCGGCTGCGCGCACGCTGTGCTTGAACACACACCGCGCGGAGCTCCG 26600
QY 249 CGCGCGCCT 257
DB 26601 CGGGCACCT 26609

RESULT 6
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA

; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 30.8%; Score 79.4; DB 9; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCCAGCAGCAGCAGTGTGCTGCTCTGAGCTGTGCGGCCACACCATGAGC 68
DB 26364 GCTGCCCGGAGTGAACACCGAGACCGGGTCCGACCTCGTCCGTACGCA 26423
QY 69 GGTGCTGAACGACGAGAAACGAGCGCACCGGTGAGTCCGGCCATCGCGAGATT 128
DB 26424 CGTGGCAGACCGTCTCGGACACG---GCAACCCGAGCCGGGTGACCTGAGAGCGGCTT 26480
QY 129 CGCCCACTCGGCTTGTGACTCCCTCATGGGTGTGCAACTGCGAACCCTCAGCAAGC 188
DB 26481 CGCGACACCGGTTTGACTCGTCAACCGCGTGAATCCGGAACCGTCTCAACCGCCG 26540
QY 189 CACGGGCTGCGGTTGCGCGTCACTCTTGAACACACACCGCGCGCGGTCCG 248
DB 26541 GACCGGCTGCGGCTGCGCGCACGCTGTGCTTGAACACACACCGCGGAGCTCCG 26600
QY 249 CGCGCGCCT 257
DB 26601 CGGGCACCT 26609

RESULT 7

US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 30.8%; Score 79.4; DB 10; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

QY 9 GCTGTCCCCCAGCAGCAGCAGTGTGCTGCTCTGAGCTGTGCGGCCACACCATGAGC 68
DB 26364 GCTGCCCGGAGTGAACACCGAGACCGGGTCCGACCTCGTCCGTACGCA 26423
QY 69 GGTGCTGAACGACGAGAAACGAGCGCACCGGTGAGTCCGGCCATCGCGAGATT 128
DB 26424 CGTGGCAGACCGTCTCGGACACG---GCAACCCGAGCCGGGTGACCTGAGAGCGGCTT 26480
QY 129 CGCCCACTCGGCTTGTGACTCCCTCATGGGTGTGCAACTGCGCAACCCTCAGCAAGC 188
DB 26481 CGCGACACCGGTTTGACTCGTCAACCGCGTGAATCCGGAACCGTCTCAACCGCCG 26540
QY 189 CACGGGCTGCGGTTGCGCGTCACTCTTGAACACACACCGCGCGCGGTCCG 248
DB 26541 GACCGGCTGCGGCTGCGCGCACGCTGTGCTTGAACACACACCGCGGAGCTCCG 26600
QY 249 CGCGCGCCT 257
DB 26601 CGGGCACCT 26609

Oy	249	CGCGCGCT	257
Db	10914	CGGCACT	10922

RESULT 2
US-09-988-384B-32
Sequence 32, Application US/09988384B
Publication No. US20030073824A1

! TYPE: DNA
! ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

```

RESULT 3
US-09-836-821-32
/ Sequence 32, Application US/09836821
/ Publication No. US20030087405A1
/ GENERAL INFORMATION:
/ APPLICANT: Sheerman, D. H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT FILING DATE: US/09/836, 821
/ PRIOR APPLICATION NUMBER: 2001-04-17
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: fastseq for Windows Version 3.0
/ SEQ ID NO: 32

```

```
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-32
```

Query Match	30.84	Score 79.4	DB 9	Length 11220
Best Local Similarity	60.24	Pred. No. 1.6e-13		
Matches 150; Conservative	0	Mismatches 96	Indels 3	Gaps 1

Db 10914 CGGGCACCCT 10922

```

! : TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

```


QY 361 GAGCT-----GACGCGGGGATGAGCTGTGCGCGTTCGCTGCTTCCTCGGCGCG 414
Db 14399 GAACCTACCTGACGCGCGGATACGACTGCGAGCGTTCCGATGTTCTCTCGCGCGCC 14458
QY 415 GGGATCTGGGGTGGCGCGCGGCACTACGCGCGCAATGCCCTCTGAGCGC 474
Db 14459 GCGCTCTTGGTGGCGCGGCGCGCGCTTACCGCGCGCAACGCGCCTCGACGCC 14518
QY 475 CTGCGCGTACCGCGCGCGCGCGCGCTGCGCGCGGGGTTGCTGCGCGCGCGCGTGG 534
Db 14519 CTGCGCTTGGCGCGCGCGCGCGCGCGCTTCTCTCGCGCTGGCGCGCTTGG 14578

RESULT 13

US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US2003007382A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,536051
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

Query Match 37.4%; Score 199.6; DB 9; Length 37948;
Best Local Similarity 62.0%; Pred. No. 3.2e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTGGTGTGTGTGAGCGGATGAGCGGATGTCTGCGTGGCGCGCGCATCTG 60
Db 15209 GGCACGCTCTGTGACCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 15268
QY 61 GCTGT 120
Db 15269 GTGGGCGAGTGGGCGGTACGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 15328
QY 121 GGTGCGGAGGATGTGTGCGGCGGAGCTGCGCGGTTGGGCGGAGGATGTTGCG 180
Db 15329 GGGCGCGAGAGTCTGTGACGAGCTGTGAGGCGCTTGGAGCGCATCTGTGTGTGT 15388
QY 181 TGTGATGTGGGAGCGCGCGGAGGTGTCTCGGCTGTGAGGAGTGTCTTGTGCGG 240
Db 15389 TGTGAGTGTGCGCGCGGAGCGCGGAGGTGTCTCGGCTGTGAGGAGTGTCTTGTG 15448
QY 241 CCGCTGACGGGATGT 300
Db 15449 CCGCTGACGGGATGT 15508
QY 301 AGCGCGGAGCGGCTGTGAGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Db 15509 AGCGCGGAGCGGCTGTGAGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 15568
QY 361 GAGCT-----GACGCGGGGATGAGCTGTGCGCGTTCGCTGCTTCCTCGCGCG 414
Db 15569 GAACCTACCTGACGCGCGCGGATACGACTTGGAGCGGCTGTGTGTGTGTGTGTGT 15628
QY 415 GGGATCTGGGGTGGCGCGCGGCACTACGCGCGCGCGCGCATGCGCGCTCTGAGCGC 474
Db 15629 GCGCTCTTGGTGGCGCGCGCGCGCGCGCGCTTCTCTCGCGCTGGCGCGCTTGG 15688

QY 475 CTGCGCGTACCGCGCGCGCGCGCGCGCTGCGCGCGGGGTTGCTGCGCGTGGCGCGCTG 534
Db 15689 CTGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTCTCGCGCTGGCGCGCTTGG 15748

RESULT 14

US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 37.4%; Score 199.6; DB 9; Length 38506;
Best Local Similarity 62.0%; Pred. No. 3.2e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTGGTGTGTGTGAGCGGATGAGCGGATGTCTGCGTGGCGCGCGCATCTG 60
Db 12181 GGCACGCTCTGTGACCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 12240
QY 61 GCTGT 120
Db 12241 GTGGGCGAGTGGGCGGTACGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 12300
QY 121 GGTGCGGAGGATGTGTGCGGCGGAGCTGCGCGGTTGGGCGGAGTGTGCGATTTGCG 180
Db 12301 GCGCGCGAGACTCTGTGACGAGCTGTGAGGCGCTTGGAGCGGACGTCTGTGTGTGT 12360
QY 181 TGTGATGTGGGAGAGCGCGGAGGTGTCTCGGCTGTGAGGAGTGTCTTGTGCGG 240
Db 12361 TGTGAGTGTGCGCGCGGAGCGGAGCGGCTTCACTGAGCGCATGCGCGCGGAGCAC 12420
QY 241 CCGCTGACGGGATGT 300
Db 12421 CCGCTGACGGGATGT 12480
QY 301 AGCGCGGAGCGGCTGTGAGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
Db 12481 AGCGCGGAGCGGCTGTGAGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12540
QY 361 GAGCT-----GACGCGGGGATGAGCTGTGCGCGTTCGCTGCTTCCTCGCGCG 414
Db 12541 GAACCTACCTGACGCGCGCGGATACGACTTGGAGCGGCTGTGTGTGTGTGTGTGT 12600
QY 415 GGGATCTGGGGTGGCGCGCGGCACTACGCGCGCGCGCGCATGCGCGCTCTGAGCGC 474


```
Db 14099 GTGGGCGAGTGGGGCGGTACGACCCCTGCTGTGTAGACCGGGGGGCAAGACGCCCCG 14158
| 121 GGTGCGAGAGGTGTGCGGGCGAGCTGGCCGCTGGGGGCGAGGTGCGGATTTGTCG 180
| 14159 GGGCGCGACGAGCTCGTGCACGAGCTGGAGAGCCCTGGAGCCGACGTCCTCGGTGGCCGG 14218
| 181 TGTGATGTGGGGAGCGGGCGGAGGTGTCCGGCTCTGAGAGGTGTTCTGCGGGGTGT 240
| 14219 TGGGAGTGTGCGCCGACCGCGAGGCCCTGACCGGCTGACGACCATCCCGCGCAACAC 14278
| 241 CCGCTGACGGGAGTGTGCTGACGCGGCTGTGTGCTGTGAGACGATGCGATGCGCTCTC 300
| 14279 CCGCTACCGCGGCTGTGCTGACGAGGAGGCTGTCTGCGACGGACCCCTCCGCTCAG 14338
| 301 AGCGCCGAGCGGCTGGGCGACGAGTGTGCGGGCCGAGGTGATGCCGCTCTTTTGTGAT 360
| 14339 ACGACGAGAGGACGTGGAAACGTAATGCGGCGCCAGGTGACGCGCGCTCTCTCGAC 14398
| 361 GAGCT-----GACGCGGGGTATGAGCTGTGCGCGCTGTGCTGTCTGCGGCGCG 414
| 14399 GAATCACTCTGACCGCCCGATGACACTGCGACGCTTCTGTCAATGTTCTCTCGCGCG 14458
| 415 GGGATCCTGGGGGTGCGCGGGCGAGGCACTACGCGCGGCAATGCGCTCTGACGCG 474
| 14459 GCGCTCTTGTGGTGGCGGGGCGAGGCGCTTACGCGCGCCCAACGCACTCTGACGCG 14518
| 475 CTGGCGTACCGGCGCGCGCGGCGGCTGTGCGGGGGGTGTGCTGCGGTGGGGGCTGTG 534
| 14519 CTGCGCTGTGCGCGCGCGGCGAGCGGACCTCCCGCCTCTCTCGGCTGGGGGCTGTG 14578
```

RESULT 11

```
US-09-836-821-5
/ Sequence 5, Application US/98836821
/ Publication No. US2003087405A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/836,821
/ PRIORITY FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: 09/105,537
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: PasteSeq for Windows Version 3.0
/ SEQ ID NO 5
/ LENGTH: 36778
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
/ US-09-836-821-5
```

Query Match 37.4%; Score 199.6; DB 9; Length 36778;

Best Local Similarity 62.0%; Pred. No. 3.2e-39;

Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

```
QY 1 GGGTCGCTGTGTGTAAGGAGTGGAGCGGATGTGCTGGAGTGGCGGAGTGGCGGACATCTG 60
| 14039 GGCACGCTCTCTGTCGACCGCGGACCGGCGGCTGGGGCGAGTGTGCGCGGCACTG 14098
| 61 GCTGTGTGTGTGGGGTGGCGGATCTGCTGTGTGTGTAAGCGCGGCTGTGCGGATCTCG 120
| 14099 GTGGGCGAGTGGGGCGTACGACGCTGTGCTGTGTGAGCGGCGGCGGCGGACGCGCC 14158
| 121 GGTGCGAGAGGTGTGCGGGGCGAGCTGGCGCGTGTGGGGCGAGGCGGATTTGTCG 180
| 14159 GGGCGCGACGAGCTGTGTCACGAGCTGGAGGCTTGGGAGCGACGTCCTCGGTGGCCGG 14218
| 181 TGTGATGTGGGGAGCGGGCGGAGGTGTGCGGCTGTGAGAGGTGTTCTGCGGGGTGT 240
| 14339 ACGACGAGAGGACGTGGAAACGTAATGCGGCGCCAGGTGAGGCGCGCTCTCTCGAC 14398
```

```
Db 14219 TGGAGCTGTGCGGACCGGAGAGCCCTGACCGCGGTACTGAGACGCAATCCCGCGAACAC 14278
| 241 CCGCTGAGGGGTGTGTGATGCGGCTGTGTGTGTGACAGAGCAATGCAATGCGCTCTCTC 300
| 14279 CCGCTACCGCGGCTGTGCTGACACGCGAGGCGCTCTCTCTGAGGAGCACTCCGCTCAG 14338
| 301 AGCGCCGAGCGGCTGGGCGACGAGTGTGCGGGCCGAGGTGATGCGGCTCTTTTGTGAT 360
| 14339 ACGACGAGAGGACGTGAAACGTAATGCGGCGCCAGGTGACGCGCGCTCTCTCTGAC 14398
| 361 GAGCT-----GACGCGGGGTATGAGCTGTGCGGGTGTGCTGTCTCTCTGCGCGCG 414
| 14399 GAATCACTCTGACCGCCCGATGACACTGAGAGGCTGTGATGTTCTCTCTCGCGCG 14458
| 415 GGGATCCTGGGGTGTGCGCGGGCGAGGCACTACGCGCGGCGCAATGCGGCTCTGACGCG 474
| 14459 GCGCTCTTGTGCTGGGCGGGGCGAGGCGCTGTACGCGCGCGCAACGCACTCTGACGCG 14518
| 475 CTGGCGTACCGGCGCGGCGGAGTGTGCGGGGGTGTGCGCTGTGAGGAGGCTGTG 534
| 14519 CTGCGCTGTGCGCGCGCGGCGAGCGGACCTCCCGCCTCTCTCTCGGCTGGGGGCTGTG 14578
```

RESULT 12

```
US-09-861-289-5
/ Sequence 5, Application US/09861289
/ Patent No. US20020110897A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/861,289
/ PRIORITY FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: 09/105,537
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: PasteSeq for Windows Version 3.0
/ SEQ ID NO 5
/ LENGTH: 36778
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
/ US-09-861-289-5
```

Query Match 37.4%; Score 199.6; DB 10; Length 36778;

Best Local Similarity 62.0%; Pred. No. 3.2e-39;

Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

```
QY 1 GGGTCGCTGTGTGTAAGGAGTGGAGCGGATGTGCTGGAGTGGCGGAGTGGCGGACATCTG 60
| 14039 GGCACGCTCTCTGTCGACCGCGGACCGGCGGCTGGGGCGAGTGTGCGCGGCACTG 14098
| 61 GCTGTGTGTGTGGGGTGGCGGATCTGCTGTGTGTGTAAGCGCGGCTGTGCGGATCTCG 120
| 14099 GTGGGCGAGTGGGGCGTACGACGCTGTGCTGTGTGAGCGGCGGCGGACGACGCGCCG 14158
| 121 GGTGCGAGAGGTGTGCGGGGCGAGCTGGCGCGTGTGGGGCGAGGTGCGGATTTGTCG 180
| 14159 GGGCGCGAGAGCTGTGTCACGAGCTGAGAGGCTTGGGAGCGGACGACGTCCTCGGTG 14218
| 181 TGTGATGTGGGGAGCGGGCGGAGGTGTCCGGCTGTGAGAGGTGTTCTCTCCGGGTGT 240
| 14219 TGGGAGTGTGCGGACCGCGAGCCCTGACCGCGCTGTAATGAGCCATCCCGCGAACAC 14278
| 241 CCGCTGACGGGAGTGTGCTGATGCGGCTGTGTGTGCTGAGACGATGCGAGATGCGCTCTC 300
| 14279 CCGCTACCGCGGCTGTGCTGACACGAGGAGGCTCTCTCTGAGGAGCACTCCGCTCAG 14338
| 301 AGCGCCGAGCGGCTGGGCGACGAGTGTGCGGGCCGAGGTGATGCGGCTCTTTTGTGAT 360
| 14339 ACGACGAGAGGACGTGGAAACGTAATGCGGCGCCAGGTGAGGCGCGCTCTCTCTGAC 14398
```

PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-836-821-30

Query Match 37.4%; Score 199.6; DB 9; Length 13842;
Best Local Similarity 62.0%; Pred. No. 3.6e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTCGCTGTTGGTGAACGGGTGGGACCGGCTGCTGGGTGCGGCGTGGCGGACATCTG 60
DB 12298 GGCACGGTCTGCTGACCGGCGGACCGGCGGCTGGGCGGACCTGGTGGCGGACGCTG 12357
QY 61 GCTGCTGTGTGTGGGGGTGGGAGATCTGCTGTTGGTGAACCGGCGTGGTCCGGATGCTCG 120
DB 12358 GTGGGGGAGTGGGGCGGCTGACACGCGCTGCTGCTGAGCGGCGGCGGACGAGACGCCCG 12417
QY 121 GGTGCGGAGGGTGTGGGGGCGGAGCTGGGCGGCGGCTGGGCGGCGGAGGAGGCGGATTTGGCG 180
DB 12418 GGGCGCGACGAGCTCTGTCACGAGCTGGAGGCGCTGGGAGCGGACGCTGCTGGCGCGG 12477
QY 181 TGTGATGTGGGGGAGCGGCGGAGGAGTGTCCGCTGCTGGAGGGTGTCTGCGCGGGTGT 240
DB 12478 TGGGAGTGGCGGACCGGCGGAGGCGGCTGACCGCGGTACTGACGCGACCCCGCGAAGAC 12537
QY 241 CGGCTGACGGGTTCTGTCATGCGGCTGGTGTGTGCTGAGCGATCGACGATGCGCTCTTC 300
DB 12538 CGGCTACCGCGGCTGTGTCACGAGGCGGCGGCTCTCTCCGACGAGCCCTCCGCTCATG 12597
QY 301 AGCGCGAGCGGCTGGGCGACAGTGTTCGCGGCGGAGGTGATCCGCTCTTTGCTGGAT 360
DB 12598 AGACGAGGAGAGCTGGAACAGTACTGGGCGGACCGGCTGACCGCGGCTCTCTGAC 12657
QY 361 GAGCT-----GACGCGGGGTATGAGCTGTGCGGCTGTGCTGTTCTCTCGCGCGG 414
DB 12658 GAACCTACGCTGACGCGCGCGGATGACGCTGGAGCGTTCATGTTCTCTCGCGCGG 12717
QY 415 GGGATCTGGGGTGGCGCGGCGGAGGCACTACGCGGCGGCAATGCGGCTCTGAGCGG 474
DB 12718 GCGCTTTCGTTGCGGCGCGGCGGCGGCGGCTTACGCGCGGCAACCGACCTCTGAGCGG 12777
QY 475 CTGGCGTACCGGCGGCGGCGGCGGCGGCTGCGCGGCGGCTGCGGCGGCGGCTGTGG 534
DB 12778 CTGGCGTGGCGCGGCGGCGGCGGCGGAGCTCCCGCTCTCTGCGGCTGGGCGCTGTGG 12837

RESULT 9

US-09-861-289-30
Sequence 30, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/861,289
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 37.4%; Score 199.6; DB 10; Length 13842;
Best Local Similarity 62.0%; Pred. No. 3.6e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTCGCTGTTGGTGAACGGGTGGGACCGGCTGCTGGGTGCGGCGTGGCGGACATCTG 60
DB 12298 GGCACGGTCTGCTGACCGGCGGACCGGCGGCTGGGCGGACCTGGTGGCGGACGCTG 12357
QY 61 GCTGCTGTGTGTGGGGGTGGGAGATCTGCTGTTGGTGAACCGGCGTGGTCCGGATGCTCG 120
DB 12358 GTGGGGGAGTGGGGCGGCTGACACGCGCTGCTGCTGAGCGGCGGCGGACGAGACGCCCG 12417
QY 121 GGTGCGGAGGGTGTGGGGGCGGAGCTGGCGCGGCTGGGCGGCGGAGGAGGCGGATTTGGCG 180
DB 12418 GGGCGCGACGAGCTCTGTCACGAGCTGGAGGCGCTGGGAGCGGACGCTGCTGGTGGCGG 12477
QY 181 TGTGATGTGGGGGAGCGGCGGAGTGTTCGCGCTGCTGAGAGGTTCTCTGCGCGGATGT 240
DB 12478 TGGGAGTGGCGGACCGGCGGAGGCGGCTGACCGCGGCTGCTGACGCGCAATCCCGCGAAGAC 12537
QY 241 CGGCTGACGGGTTCTGTCATGCGGCTGGTGTGTGCTGAGCGATGCGATGCGCTCTTC 300
DB 12538 CGGCTACCGCGGCTGTGTCACGAGGCGGCGGCTCTCTGACGCGACCTCTCCGCTCATG 12597
QY 301 AGCGCGAGCGGCTGGGCGACAGTGTTCGCGGCGGAGGTGATCCGCTCTTTGCTGGAT 360
DB 12598 AGACGAGGAGAGTGGGAACAGTACTGGGCGGACCGGCTGACCGCGGCTCTCTGAC 12657
QY 361 GAGCT-----GACGCGGGGTATGAGCTGTGCGGCTGTGCTGTTCTCTCGCGCGG 414
DB 12658 GAACCTACGCTGACGCGCGCGGATGACGCTGGAGCGTTCATGTTCTCTCGCGCGG 12717
QY 415 GGGATCTGGGGTGGCGCGGCGGAGGCACTACGCGGCGGCAATGCGGCTCTGAGCGG 474
DB 12718 GCGCTTTCGTTGCGGCGGCGGCGGCGGCGGCTTACGCGCGGCAACCGACCTCTGAGCGG 12777
QY 475 CTGGCGTACCGGCGGCGGCGGCGGCGGCTGCGCGGCGGCTGCGGCGGCGGCTGTGG 534
DB 12778 CTGGCGTGGCGCGGCGGCGGCGGCGGAGCTCCCGCTCTCTGCGGCTGGGCGCTGTGG 12837

RESULT 10

US-09-860-846-5
Sequence 5, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 36778
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 37.4%; Score 199.6; DB 9; Length 36778;
Best Local Similarity 62.0%; Pred. No. 3.2e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTCGCTGTTGGTGAACGGGTGGGACCGGCTGCTGGGTGCGGCGTGGCGGACATCTG 60
DB 14039 GGCACGGTCTGCTGACCGGCGGACCGGCGGCTGGGCGGAGCTGTGCGCGGACGCTG 14098
QY 61 GCTGCTGTGTGTGGGGGTGGGAGATCTGCTGTTGGTGAACCGGCGGCTGGATGCTCGG 120

QY 481 TACCGGCGCGGCGGCGGCTGCGCGGCGGCTGCGCGGCGGCTGCG 534
 Db 7817 GCCACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCG 7870

RESULT 4

US-09-861-289-1
 ; Sequence 1, Application US/09861289
 ; Patent No. US2002010897A1
 ; GENERAL INFORMATION
 ; APPLICANT: Sherman, D.H.
 ; APPLICANT: Liu, H.
 ; APPLICANT: Xue, Y.
 ; APPLICANT: Zhao, L.
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
 ; FILE REFERENCE: 600,438US1
 ; CURRENT APPLICATION NUMBER: US/09/861,289
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: 09/105,537
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 15872
 ; TYPE: DNA
 ; ORGANISM: Streptomyces venezuelae
 ; US-09-861-289-1

Query Match 39.8%; Score 212.4; DB 10; Length 15872;
 Best Local Similarity 62.4%; Pred. No. 3,1e-42;
 Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTGTAACGAGGTGGAGCGGCTGTGCTGGGTGGCGCGGATCTG 60
 Db 7337 GGAACCGTCTCTCTCAACCGGCGGCGACCGAGGCTCGAGGCTCTCTCGCCCGACCTC 7396
 QY 61 GCTGTGTGTGTGGGTGGCGGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
 Db 7397 GTTGAACGTTCACCGGCGGCGGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7456
 QY 121 GGTGCGGAGGCTGTGTGGGCGGAGCTGTGGCGGCTGTGGGCGGAGGTGGCGGATTTGTTGG 180
 Db 7457 GGGCGGCGGACCTGTGGCGGAGACCTGTGGCGGCGGCGGCGGAGGTGTGTGTGTGTGTGT 7516
 QY 181 TGTGATGT 240
 Db 7517 GCGGAGCGGCGGCGGCGGAGACCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7576
 QY 241 CCGCTGACGAGGT 300
 Db 7577 CCGCTGACGAGGT 7636
 QY 301 AGCGCGGAGCGGCTGTGGGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
 Db 7637 AACCGGAGCGGCTGT 7696
 QY 361 GAGCTGACGCGGCGGATGT 420
 Db 7697 GAGCTGACGAGGACCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7756
 QY 421 CTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 480
 Db 7757 GTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 7816
 QY 481 TACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 534
 Db 7817 GCCACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 7870

RESULT 5
 US-09-712-363-116
 ; Sequence 116, Application US/09712363
 ; Patent No. US20020164588A1

; GENERAL INFORMATION:
 ; APPLICANT: Eisenberg, David
 ; APPLICANT: Rotstein, Sergio H.
 ; APPLICANT: Marcotte, Edward M.
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 ; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
 ; FILE REFERENCE: 07419-032001
 ; CURRENT APPLICATION NUMBER: US/09/712,363
 ; PRIOR FILING DATE: 2000-11-13
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246
 ; PRIOR FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/179,531
 ; PRIOR FILING DATE: 2000-02-01
 ; PRIOR APPLICATION NUMBER: 60/117,844
 ; PRIOR FILING DATE: 1999-01-29
 ; PRIOR APPLICATION NUMBER: 60/118,206
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: 60/126,593
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 60/134,093
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/134,092
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/165,124
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: 60/165,086
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 292
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 116
 ; LENGTH: 4851
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-712-363-116

Query Match 38.8%; Score 207.4; DB 9; Length 4851;
 Best Local Similarity 62.9%; Pred. No. 5,5e-41;
 Matches 338; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

QY 1 GGGTCGGT 60
 Db 3682 GGCACGCTGTGTATCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3741
 QY 61 GCTGT 120
 Db 3742 GTTGGCGCTATGGGT 3801
 QY 121 GGTGCGGAGGCTGT 180
 Db 3802 GAGCGGCGGCGGATTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3861
 QY 181 TGTGATGT 240
 Db 3862 TGTGATGT 3921
 QY 241 CCGC---TGAAGGCTGT 297
 Db 3922 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3981
 QY 298 CTCACCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
 Db 3982 TTGAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4041
 QY 358 GATGAGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 417
 Db 4042 CACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4101
 QY 418 ATCTGAGGCTGT 477
 Db 4102 ACGGTGCTGT 4161
 QY 478 GCGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 534

[illegible]

RESULT 2

```

1 Sequence 1, Application US/09988384B
2 Publication NO. US20030073824A1
3 GENERAL INFORMATION:
4 APPLICANT: Sherman, D. H.
5 APPLICANT: Liu, H.
6 APPLICANT: Xue, Y.
7 APPLICANT: Zhao, L.
8 TITLE OF INVENTION: DNA encoding methymycin and plikomycin
9 FILE REFERENCE: 600,536US1
10 CURRENT APPLICATION NUMBER: US/09/988,384B
11 CURRENT FILING DATE: 2001-11-19
12 PRIOR APPLICATION NUMBER: PCT/US99/14398
13 PRIOR FILING DATE: 1999-06-25
14 PRIOR APPLICATION NUMBER: US 09/105,537
15 PRIOR FILING DATE: 1998-06-26
16 NUMBER OF SEQ ID NOS: 53
17 SEQ ID NO: 1
18 LENGTH: 15872
19 TYPE: DNA
20 ORGANISM: Streptomyces venezuelae
21 US-09-988-384B-1

```

Query Match

Query Match	39.8%	Score 212.4	DB 9	Length 15872
Best Local Similarity	62.4%	Pred. No. 3.1e-42		
Matches 333, Conservative	0	Mismatches 201	Indels 0	Gaps 0

Qy	1	GGGTCCGGTGTGGTGA	CCGGSTGGGA	CCGGGTGAG	CTCGGGTAC	CGGCGTGG	CCGGCATATG	60
Db	7337	GGACACCGTCTCT	CGTCA	CCGGGGGCA	CCGGAGGCTCT	CGACCCGCA	CTTC	7336
Qy	61	GCTGTGTGTGTGGG	ATCGGGGAT	CTGTTTGT	TGA	CCGGAGTGT	CCGAGTCTCG	120
Db	7397	GTGGACCGTTC	ACGGGGGTGGCC	CGGGCTGT	GTGTGAG	CCGGCGGGGCG	CAACCCCC	7456
Qy	121	GGTGGCGGAGG	GTCTGGGGCGGAG	CTGGG	CCGGTTTGGGGGGCGG	AGATGGATTTGTG	CG	180
Db	7457	GGCGCGGCGCA	CTGGGGCGAGAC	CTTCGGGGCT	TCGGGCGG	AGATGGATTTGTG	CG	7516
Qy	181	TGTGATGTGGGG	AGCGCGGAGGTGT	CTCGGCTGT	GTGAAGG	TGTTCTGC	CGGGTGT	240
Db	7517	GCGGACGCC	CCGACCGCGAG	ACCTTGGGCGGGCG	ATGTCCAC	CGTGTCCGCG	CCAGAT	7578
Qy	241	CCGCTGACGGG	GTCTGTGATCGG	CTGTGTGTCTGA	CGATGCA	CAATCGCTCT	CTC	300
Db	7577	CCCTGACCGG	CGCTGTGTGAC	ACGGGGAGTGTCTGA	CGACGCA	CGTGTGA	GGCGCTC	7636
Qy	301	ACGCCGAGCGG	CTGGGCA	CGGTGTTCCG	GGCCAGAGTGG	ATGCGGCTCTTTTGT	CGAT	360
Db	7637	ACACCGGAA	CGGCTGGACCGGGTACT	GGCCCGGAG	GTGAGCGCGCGCTGTGA	ACTTGGAC		7698

Qy 361 GAGCTGACGCGGGGAGATAGAGACTGACGCGGTTGCGCGTCTCCTCGGCGCGGGAGATC 420

Db 7697 GAGCTCAACAAAGAGACTCGGCTCGACGCGCTTCGTCCTCTCTCCTCGGCTTCGCGATC 775

Qy 421 CTGGGGGTGCGCGGCGCAGGAGCACTACGCGCGGCGCAATGCGGCTCTGGACGCGGTGGCG 480

Db 7757 GTGGGCACTCGCGGCGAGGCGCAACTACGCGGCGGCGCAACACGCGGCTTCGACGCGCTTCGCC 781

Qy 481 TACCGGCGGCGGCGCGCGGCTCTGCGCGGGGAGTGCCTGCGCGTGGGCGCTTGG 534

Db 7817 GCCCACGCGCGCGCGACGCGGCTTCGCGCGGCTTCGCGCGCTTCGCGCGCTTCGG 7870

RESULT 3

```

US-09-836-821-1
; Sequence 1, Application US/09636821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-1

```

Query Match

Query Match	39.8%	Score 212.4	DB 9	Length 15872
Best Local Similarity	62.4%	Pred. No. 3.1e-42		
Matches 333, Conservative	0	Mismatches 201	Indels 0	Gaps 0

QY	1	GGAGTCGGATGTTGGATGACCGGGTGGGACCGGGATGCTCGGGATGCGCGCGGACATCTG	60
Db	7337	GGCACCGTCTCTGTCACACCGGGGGGACCGGAGAGCTTCGAGAGCGCTTCCTGCGCCGCACTCT	7336
QY	61	GCTGATGTGTGTGGAGGTGCGGGATCTGCTGTGTGTGAGCCGGGCGTGTCTCGGATCTCCG	120
Db	7397	GTGAGCGCTACCGGGGTGTCCGCGCGGCTGTGCTGTGTGAGCGCGCGGGGGCGGACGCCCC	7456
QY	121	CGTCCGGAGAGGTCTTCCGGGCGGAGCTGGCCGCGCTTGGGGGCGGAGAGTGTCCGGATTGTTCG	180
Db	7457	GGCCCGGCGGACCTGTGGCGGAGAGACCTGCGCGGGCCTTCGCGCGGAGAGTGTGCTTGCGCCC	7516
QY	181	TGTGATGTGTGGGGAGCGCGCGGGAGGTGGTTCGGCTGCTGAGAGGTGTTCTTCGCGGGGT	240
Db	7517	GCCAGCGCCCGGACCGCGGAGAGGCTGGCCGCGGGCGATCGCCACCTGTCCGCCGAGCAT	7576
QY	241	CCGCTGACGGGTGTCTGTGCATGCGGCTGTGTGTGCTTGACGATGTGCAGCATGCTTCTTC	300
Db	7577	CCGCTGACGGCGCTGTGTGTGCACACCGGCGGAGTCTGTGCACGACGAGACGGTGTGAGGCGCTC	7636
QY	301	ACGCCCGGACGGGTGTGGGACAGGTGTTCGGGGCCCAAGTGTGATGTGCGCTCTTTTCTGGAT	360
Db	7637	ACACCGGAAACGGCTGTGAGCGGTACTGTGCGCCCGGAAGGTGTGACGCGCGCGGTGAACTGTGAC	7696
QY	361	GAGCTGACGCGGGGTATGTGAGCTGTTCGCGCTTCTGTGCTGTCTCTCGGCGCGGGGATC	420
Db	7697	GAGCTCACACAGGACCTGTGGGCTGTGAGCGCTTGTGTCTCTCTCTCGCTTCGGGATC	7756
QY	421	CTGGGGTGTGCGCGGACGGGCACTACGCGCGGCGCAATGTCCGCTGTGATGCGGTGGCG	480
Db	7757	GTCCGCAACCGCTGGCGCAGGCCCACTTACCGCGCGGCGCAACAGGGGCTTGTGAGGCCCTTCGC	7816

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 63.1072 Seconds
(Without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142

Perfect score: 534

Sequence: 1 ggggtcggtgtgtgtgacggg.....cgctgcgtcggtgtgtg 534

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.4	39.8	15872	9 US-09-860-846-1	Sequence 1, Appl1
2	212.4	39.8	15872	9 US-09-988-384B-1	Sequence 1, Appl1
3	212.4	39.8	15872	9 US-09-836-821-1	Sequence 1, Appl1
4	212.4	39.8	15872	10 US-09-861-289-1	Sequence 1, Appl1
5	207.4	38.8	14851	9 US-09-712-363-116	Sequence 116, App
6	199.6	37.4	13842	9 US-09-860-846-30	Sequence 30, Appl1
7	199.6	37.4	13842	9 US-09-988-384B-30	Sequence 30, Appl1
8	199.6	37.4	13842	9 US-09-836-821-30	Sequence 30, Appl1
9	199.6	37.4	13842	10 US-09-861-289-30	Sequence 30, Appl1
10	199.6	37.4	36778	9 US-09-860-846-5	Sequence 5, Appl1
11	199.6	37.4	36778	9 US-09-836-821-5	Sequence 5, Appl1
12	199.6	37.4	36778	10 US-09-861-289-5	Sequence 5, Appl1
13	199.6	37.4	37948	9 US-09-988-384B-5	Sequence 5, Appl1
14	199.6	37.4	38506	9 US-09-793-708-19	Sequence 19, Appl1
15	196.4	36.8	11220	9 US-09-860-846-32	Sequence 32, Appl1
16	196.4	36.8	11220	9 US-09-988-384B-32	Sequence 32, Appl1
17	196.4	36.8	11220	9 US-09-836-821-32	Sequence 32, Appl1
18	196.4	36.8	11220	10 US-09-861-289-32	Sequence 32, Appl1
19	183.6	34.4	68750	9 US-10-014-717-1	Sequence 1, Appl1

20	166	31.1	50937	9 US-09-808-880-1	Sequence 1, Appl1
21	161.8	30.3	552	9 US-09-957-483-32	Sequence 32, Appl1
22	157	29.4	552	9 US-09-957-483-34	Sequence 34, Appl1
23	132	24.7	4689	9 US-09-860-846-34	Sequence 34, Appl1
24	132	24.7	4689	9 US-09-988-384B-34	Sequence 34, Appl1
25	132	24.7	4689	9 US-09-836-821-34	Sequence 34, Appl1
26	132	24.7	4689	10 US-09-861-289-34	Sequence 34, Appl1
27	87.6	16.4	12793	9 US-10-032-393-8	Sequence 8, Appl1
28	87.6	16.4	12793	9 US-09-864-761-20241	Sequence 20241, A
29	83.8	15.7	1635	10 US-09-864-761-3471	Sequence 3471, Ap
30	83.8	15.7	1973	10 US-09-864-761-19531	Sequence 19531, A
31	80.8	15.1	5484	9 US-09-712-363-115	Sequence 115, App
32	79.8	14.9	1075	10 US-09-864-761-15241	Sequence 15241, A
33	79.8	14.9	1403	10 US-09-864-761-2513	Sequence 2513, Ap
34	74.4	13.9	446	10 US-09-864-761-20699	Sequence 20699, A
35	70.6	13.2	504	10 US-09-878-574-4330	Sequence 4330, Ap
36	66.8	12.5	152331	9 US-10-095-407-16	Sequence 16, Appl1
37	66.6	12.5	830	10 US-09-864-761-19531	Sequence 19531, A
38	65.4	12.2	768	9 US-09-938-842A-812	Sequence 812, App
39	64.2	12.0	2283	9 US-09-938-842A-1691	Sequence 1691, Ap
40	64.2	12.0	2283	10 US-09-887-576-645	Sequence 645, App
41	62.6	11.7	4339	9 US-10-171-311-157	Sequence 157, App
42	62.6	11.7	4339	10 US-09-964-824A-105	Sequence 105, App
43	62.6	11.7	4339	10 US-09-964-824A-578	Sequence 578, App
44	62.6	11.7	4339	10 US-09-864-864-354	Sequence 354, App
45	62.6	11.7	4339	10 US-09-880-107-2121	Sequence 2121, Ap

ALIGNMENTS

RESULT 1
US-09-860-846-1
Sequence 1, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438051
CURRENT APPLICATION NUMBER: US/09/860,846
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 39.8%; Score 212.4; DB 9; Length 15872;
Best Local Similarity 62.4%; Pred. No. 3.1e-42;
Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY	1	GGGTCGTTGTGTGTGACGGGTGGACGGGTGCTGGGTGGCGGCGGATCTCGG	60
DB	7337	GGACACGTCCTCTGTACACCGGGGACACCGAGGCTCTCGGAGCCGACCTTC	7396
QY	61	GCTGCTGTGTGGGGTGGCGGATCTGCTGTTGGTGGAGCGGCGGATGCTCGG	120
DB	7397	GTGAGACCTTCACGGGGTGGCGGCTGCTGTTGGTGGAGCGGCGGATGCTCGG	7456
QY	121	GGTTCGAGAGCTCTGCGCGGCGGAGCTGCGGCTTGGGGCGGAGTGGATTTGGC	180
DB	7457	GGGCGGCGGACGACTGGGCGAGGACCTCGGGGCTCGGGCGGAGTGGCGTTCC	7516
QY	181	TGTGATGTGGGAGCGGCGGAGGTTGCTCGGCTGCTGAGAGTGTCTTCTCGG	240
DB	7517	GCCGACGCGCGGACGCGAGGAGCTGCGCGGGGAGTGCACCGTGCCTCCGCG	7576

THIS PAGE BLANK (USPTO)


```
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
; US-08-804-227C-13

Query Match      37.1%; Score 198; DB 2; Length 13987;
Best Local Similarity 62.0%; Pred. No. 2e-29;
Matches 331; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

QY      1  GGTCGGTGTGTGAGCGGTGGGACGGGTGCTGGGTGCGCGCGGTGGCGGCATCTG 60
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7034  GGACACGTACTGATCACCGGGGCACCGGTGCGCTGGGTGCGCAGGTGCGCCGAGGCTC 7093

QY      61  GCTGGGTGTGTGGGTGCGGGATCTGCTTTGGTGAGCGCGGTGCTCCGGATGCTCG 120
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7094  GCCCGGTGCGGCCCGCGCGC---TCTGCTCTGTTGGGCGCGCGCGCGCGCGCCCC 7150

QY      121  GGTGCGAGGGTCTGCGGCGGAGCTGCGCGCTTGGGGCGGAGGTGCGGATTGTTGCG 180
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7151  GGAGTGGCGAATCTCTGAGGAGCTGACGGCGCTCGTTCCGAAGTGGCCGTCGAGGCC 7210

QY      181  TGTGATGTGGGGAGCGCGGGAGGTGTTCGGCTCTCTGAGAGGTGTTCTTCCCGGTGT 240
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7211  TGGAGCTGCGCCACCGGACGCACTGGCGCGCTCTCTCGCGGCTTCCCGAGAGCGG 7270

QY      241  CCGCTGACGGGTCTCGTGCATGCGCTGGTGTGCTGGACGATGCGACGATCGCCTCTC 300
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7271  CCCCTCGTCCCGTACTGACCGGCGAGGTGTCTGACGAGGTGTGTCTGACTCGCTC 7330

QY      301  ACGCCCGAGCGGTGGGCACTGTTTCGCGGCCAAGTGGATGCCCTCTTTTGTGGAT 360
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7331  ACCTCCGACCGGTGGACGCCGTACTCGGGACAGGTCAACCGCGCCGTCACCTGGAC 7390

QY      361  GAGCTGACCGGGGTATGGAGCTGTGCGGTTCGTCTGTCTCTCGCGCGCGGGATC 420
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7391  GAGCTGACCGCGACCTTCGCTCGACGCTTGTGCTCTTCTCTCCATCGTCGCGGTG 7450

QY      421  CTGGGTTCGCGCGGCGAGGCAACTACGCGCGGCCAATGCCGCTCTGACCGGCTGGCG 480
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7451  TGGGGCAACGGAGGCGAGGCCGTCTACCGGCGGCCAACCGCGCGCTCGACGCCCTGGCG 7510

QY      481  TACCGCGCGCGCGCGGCTCTGCGGGGTGTCTGCGGTGGGGGCTGTGG 534
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7511  CAGCGCGCGCGGCGAGGGAGCCCGTGCCTGATCGCTGGGGGCGGTGG 7564
```

Search completed: June 18, 2003, 01:08:14
Job time : 47.1323 secs

Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;
Qy 1 GGGTCGGTGTGGTACCGGGTGGGACGGGTGTGCTGGGTGGGCGGTGGCGGCATCTG 60
Db 14039 GGCACGGTCTCTGTACCGCGGGCACCGCGGCTTGGCGGACTGGTGGCGCGCACGTG 14098
Qy 61 GCTGCTGTGTGGGGTGGGATCTGCTGTGGTGGAGCGCGGTGGTCCGGATGCTCCG 120
Db 14099 GTGGCGAGTGGGGGTACGACGCTGTGCTGGTGGAGCGCGGTGGTCCGGATGCTCCG 14158
Qy 121 GGTGCGAGGGTCTCGGCGGAGTGGCGCGTGGTGGGGGGAGGTGGGATGTTGCG 180
Db 14159 GCGCGCGACGAGTCTGTGACGAGTGGAGGCTTGGAGCGGCTGTGCTGGTGGCGCG 14218
Qy 181 TGTGATGTGGGGAGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTCTTGGCGGGTGT 240
Db 14219 TCGCAGCTCGCGGACCGGAGCCCTCACCGCGTACTCGAGCGCATCCCGCGGAAAC 14278
Qy 241 CGCTGACGGGTGTGTCATCGCGTGTGTGCTGGAGCATGGACGATCGCCTCTCTC 300
Db 14279 CGCTCACCGGTGTGTCACAGGAGGCTCTCTCGAGCGGACCTCCCGTCCATG 14338
Qy 301 ACGCCGAGCGCTGGGACAGTGTTCGGCGCAAGGTGGATGCGCGTCTTTTGTGAT 360
Db 14339 ACGACGGAGGAGTGGAAACACACTACTGCGGCCCAAGGTGAGCGCGGTCTCTCGAC 14398
Qy 361 GAGCT-----GACGGGGTATGAGCTGTGGGGTTCGCTGTTCTTCTCGCGCGG 414
Db 14399 GAACTCACCTGACGCGCGGCATACGACCTGGAGCGTTCGTGATGTTCTCTCGCGCGC 14458
Qy 415 GGGATCTTGGGTTCGGCGGCGGAGGCACTACGCGCGGCGCAATGCGCTCTGGACGG 474
Db 14459 GCGCTTTCGGTGGCGGGGAGGGCGCTACGCGCGCGCAAGCCCTCGACGCC 14518
Qy 475 CTGGGTACCGCGCGGCGCGGTCTGCGGGGGTGTGCTGGCGGGTGTGGGGTGTGG 534
Db 14519 CTGGCTGGCGCGCGGCGGAGCTCCCGCGCTCTCTCGGTGGGGCTCTGG 14578

RESULT 14

US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 3062202120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 37.4%; Score 199.6; DB 3; Length 38506;
Best Local Similarity 62.0%; Pred. No. 9.9e-30;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;
Qy 1 GGGTCGGTGTGGTACCGGGTGGGACGGGTGTGCTGGGTGGGCGGTGGCGGCATCTG 60
Db 12181 GGCACGGTCTCTGTACCGCGGGCACCGCGGCTTGGCGGACTGGTGGCGCGCACGTG 12240
Qy 61 GCTGCTGTGTGGGGTGGGATCTGCTGTGGTGGAGCGCGGTGGTCCGGATGCTCCG 120
Db 12241 GTGGCGAGTGGGCGTACGACGCTGTGCTGGTGGAGCGCGGTGGGACGCGCGCG 12300
Qy 121 GGTGCGAGGGTCTCGGCGGAGTGGCGCGTGGGGGGAGGTGGGATGTTGCG 180
Db 12301 GCGCGCGACGAGTCTGTGACGAGCTGGAGGCCCTGGGAGCGGACGTCTCGTGGCGCG 12360
Qy 181 TGTGATGTGGGGAGCGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTCTTGGCGGGTGT 240
Db 12361 TGGACGCTGCGCGACCGGAGCCCTCACCGCGTACTCGAGCGCATCCCGCGGAAAC 12420
Qy 241 CGCTGACGGGTGTGTCATCGCGTGTGTGCTGGAGCATGGACGATCGCCTCTCTC 300
Db 12421 CGCTCACCGCGTGTGTCACAGGAGGCTCTCTCGAGCGGACCTCCCGTCCATG 12480
Qy 301 ACGCCGAGCGCTGGGACAGCGTGTTCGGCGCAAGGTGGATGCGCGTCTTTTGTGAT 360
Db 12481 ACGACGGAGGAGTGGAAACACACTACTGCGGCCCAAGGTGAGCGCGGTCTCTCGAC 12540
Qy 361 GAGCT-----GACGGGGTATGAGCTGTGGGGTTCGCTGTTCTTCTCGCGCGG 414
Db 12541 GAACTCACCTGACGCGCGGCATACGACCTGGAGCGTTCGTGATGTTCTCTCGCGCGC 12600
Qy 415 GGGATCTTGGGTTCGGCGGCGGAGGCACTACGCGCGGCGCAATGCGCTCTGGACGG 474
Db 12601 GCGCTTTCGGTGGCGGGGAGGGCGCTACGCGCGCGCAAGCCCTCGACGCC 12660
Qy 475 CTGGGTACCGCGCGGCGCGGTCTGCGGGGGTGTGCTGGCGGGTGTGGGGTGTGG 534
Db 12661 CTGGCTGGCGCGCGGCGGAGCGGACTCCCGCGCTCTCTCGGTGGGGCTCTGG 12720

RESULT 15

US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:

US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 39.8%; Score 212.4; DB 4; Length 15872;
Best Local Similarity 62.4%; Pred. No. 4e-32;
Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 1 GGCTCGGTGTGTGACGGGTGGGACGGGTGTGCTGGGTGCGCGGTGGCGGCATCTG 60
DB 7337 GGCACCGTCTCTGTCACCGGGGGACCGGAGGCTCGGAGGCTCTCGCCGCCACCTC 7396

QY 61 GCTGTGTGTGTGGGTGCGGATCTGCTGTGGTGAAGCGGGGTGGTCCGATGCTCG 120
DB 7397 GTGAGCGTACCGGGGTGCGCGGTGCTGCTGTGAGCGCGCGGGCGGACGCCCC 7456

QY 121 GTGTGGAGGCTCTGCGGCGGAGCTGGCGGCTTGGGGGCGGAGTGGGATTTGG 180
DB 7457 GCGCGGCCGACCTTGGCGGAGACCTCGCGGGCTCGGCGGAGTGGCGTTGGCCGC 7516

QY 181 TGTGATGTGGGAGCGCGGAGGTGTGTCGGCTCTGCGAGGTGTTCTGCGGGGT 240
DB 7517 GCGGACGCCGCGGACCGGAGGCTTGGCGGGGATCGCCACGTCGCCCGGAGCAT 7576

QY 241 CGCTGACGGGTGTGTCATGCGGTGTGCTGTGGTGGAGTGGAGTGGCTCTCTC 300
DB 7577 CGCTGACGGGTGTGTCATGCGGTGTGCTGTGGTGGAGTGGAGTGGCTCTCTC 7636

QY 301 AGCGCGGAGCGGTGGGACGGGTGTGCGGCGCAAGTGGTGGCTCTTTTGTGGAT 360
DB 7637 ACACGGACGCTGACCGGCTGTGCGCGGAGTGGCGCGGAGTGGAGTGGCTCTC 7696

QY 361 GAGCTGACCGGGGTATGAGTGTGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 7697 GAGCTCACCAAGACCTGCGGCTCGAGCGCTTCTGCTGCTGCTGCTGCTGCTG 7756

QY 421 CTGGGGTGGCGGGGAGGGGCACTACGGCGGGGCAATGCGGCTCTGAGCGGCTGGG 480
DB 7757 GTGCGACCGCGCGGCGAGCCCACTACGCGGGGCGCAACACGCGGCTCTGAGCCCTCGCC 7816

QY 481 TACCGCGCGGCGGCGGCTGTGCGGGGGTGTGCTGCGGTGGGGGTGG 534
DB 7817 GCCCACCAGCGCGGCGGCTTGGCGGAGTGTGCTGCGGCTTGGGGCTCTGG 7870

RESULT 12
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A

; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match 37.4%; Score 199.6; DB 4; Length 13842;
Best Local Similarity 62.0%; Pred. No. 9.8e-30;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1 GGGTGGGTGTGTCACGGGTGGGACGGGTGTGCTGGGTGGCGGTGGCGGCATCTG 60
DB 12298 GGCACGGTCTCTGTCACCGGGGACCGGCGGCTTGGGCGGAGTGTGCGCGGACGTG 12357

QY 61 GCTGTGTGTGTGGGTGCGGATCTGCTGTGGTGAAGCGGCGGTGGTCCGATGCTCG 120
DB 12358 GTGGCGGAGTGTGCGGCGGAGCTGGCGGCTGCTGTGCTGTGAGCGCGGCGGACGACG 12417

QY 121 GGTGCGGAGGTCTGCGGCGGAGCTGGCGGCTTGGGGGCGGAGTGGGATTTGG 180
DB 12418 GCGCGCGACGAGCTGTCACGAGTGGAGGCCCTTGGAGCGGACGTCTCGGTGGCGCG 12477

QY 181 TGTGATGTGGGGGAGCGGCGGAGGTGTGTCGGCTGTGAGGGGTGTTCTTCTGCGGGTGT 240
DB 12478 TCGGAGTGTGCGGACCGGAGGCTTACCGGCGGTACTCGAGCCATCTCCCGCGGACAC 12537

QY 241 CGCTGACGGGTGTGTCATGCGGTGTGCTGTGGTGGAGTGGAGTGGCTCTCTC 300
DB 12538 CGCTTCTTCTGCGGTGTGTCACACGCGGCGGCTCTCGGAGGACCCCTCCGCTCATG 12597

QY 301 AGCGCGGAGCGGTGGGACCGGTGTGCGGCGCAAGTGGAGTGGCTCTTTTGTGGAT 360
DB 12598 ACAGCGGAGGACGTGGAAACACGTACTGCGGCCCAAGGTGAGCGCGGTCTCTCTCGAC 12657

QY 361 GAGCT-----GACCGGGGTATGAGCTGTGCGGCTTCTGCTGCTGCTGCTGCGCGCG 414
DB 12658 GNACTCACTCGACCGCGGATGAGCTGAGCTGAGGAGGTTGCTGATGTTCTCTCGCGCG 12717

QY 415 GGGATCTGGGTGCGCGCGGAGGCAACTACGCGCGGCGCAATGCGGCTCTGAGACGG 474
DB 12718 GCGCTTCTGCGTGGCGCGGCGGCGGCTTACGCGCGCGGCAACGCGACCCCTCGACGCC 12777

QY 475 CTGGGCTACCGCGCGGCGGCGGCTGTCGCGGGGTGTGCTGCGGTGGGGGTCTGG 534
DB 12778 CTCGCTTGGCGCGCGGCGGCGGACTCCCGCCCTCTCTCTGCGGTGGGGCTCTGG 12837

RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match 37.4%; Score 199.6; DB 4; Length 36778;
Best Local Similarity 62.0%; Pred. No. 9.8e-30;

```
QY 358 GATGAGCTGACGCGGGGTATGAGCTGTGCGCGTTCGTGCTTCTCTCCGCGCGGGG 417
Db 19901 CACGAGCTGACGCGGAGTGTGATCTCGCGCGTTCGTGCTTCTCTCCGCGCGGGT 19960
QY 418 ATCTCGGTGCGCGCGGAGGCAACTACGCGCGGCAATGCGCTCTGACGCGCTG 477
Db 19961 ACCTTTGGCGCGCGCGGAGCAACTACGCGCGGCAACACAGTTCTCTCGACGCGCTC 20020
QY 478 GCGTACCGCGCGCGCGGCTCTCGCGGGGTCTCGCGGTGCTCGCGTGGGGCTGTGG 534
Db 20021 CGCGCGACCGCGCGCGGCTCTCGCGGAGCTCGCGGAGCTGCGGTGCTCTCG 20077

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 41.0%; Score 218.8; DB 4; Length 4403765;
Best Local Similarity 63.1%; Pred. No. 2.4e-33;
Matches 337; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGAGCGGTGGACGGGTGCTGGGTGCGCGGTGGCGCGGCATCTG 60
Db 1871256 GGCACCGTGTGTCTACTGGGGGACCGGATGCGCGGTTCGCGGTGGCTACCCATTG 1871315
QY 61 GCTGGTGTGTGCGGGTGGGATCTGCTGTGAGCGCGGTGTCGGATGCTCCG 120
Db 1871316 GTCCGCGACATGGGTGGCCAACTGTTCTGTCAGCGGAAGCGGTGAGCAGGCGAC 1871375
QY 121 GGTGCGGAGGTCTGCGGGGAGCTGGCGCTGTGCGCGGTGGCGGAGTGTGCG 180
Db 1871376 AGGGCGGAGAACTGCGGCGCTGTTGCGGAGGGCGGGGCCAGGTGGCGGTCTCC 1871435
QY 181 TGTGATGTGGGAGCGGGAGGTGTCGGCTGCTGAGGGGTGTTCTCTCGCGGTGT 240
Db 1871436 TGTGATGTGCTGATCTGATGCTGCTGGCGGCAATGTTGGCGGGTCTGATCCGCGCTAT 1871495
QY 241 CCCTGACGCGGTGTCTGATGCGGGTGTGCTGAGCGATCGGATCGGCTCTCTC 300
Db 1871496 CCCTTAAAGGGGTGTTTCATGCGCTGGGGTGTGAGCGATCCGCTGATCAGGCGT 1871555
QY 301 ACGCCGAGCGGTGGGACGCGGTGTCGGCGGCAAGTGTGATCGCGCTCTTTTGTGGAT 360
Db 1871556 ACACCGGATCGGTGGATACGCGGTGTCGGGCGCAAGTGTGATGGGCGCTTGGATCTGCAC 1871615
QY 361 GAGCTGACGCGGGTATGAGCTGTGCGGTTCGTGCTTCTCTCGCGCGGGATC 420
Db 1871616 GAGCTAACGAGGACATGGAATTTGTCGGCGTTCGTGCTTCTCTCGCGCGGGAT 1871675
QY 421 CTGGGTGCGCGCGGAGGCAACTACGCGCGGCAATGCGCTCTTGGACGCGCTGGCG 480
```

```
Db 1871676 GTGGGCAACCGGCTCAGGGGAATATGCTCGCGGAATGCGTTTTCGACGGGTGCTG 1871735
QY 481 TACCGCGCGCGCGCGGTCTCCCGGGGTGTCGCTGGCGTGGGGCTGTGG 534
Db 1871736 GCCTATCGCGCTCGCGTGGGCTGCGCGAATTGTCGGTGGCGTGGGACTGTGG 1871789

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 41.0%; Score 218.8; DB 4; Length 4411529;
Best Local Similarity 63.1%; Pred. No. 2.4e-33;
Matches 337; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGAGCGGTGGACGGGTGTCGGGTGCGCGGTGGCGCGGCATCTG 60
Db 1880381 GGCACCGTGTGTCTACTGGGGGACCGGCAATGCGCGGTTCGCGGTGGCTACCCATTG 1880440
QY 61 GCTGGTGTGTGCGGGTGGGATCTGCTGTGAGCGCGGTGTCGGATGCTCCG 120
Db 1880441 GTCCGCGACATGGGTGGCCAACTGTTCTGTCAGCGGAAGCGGTGAGCAGGCGAC 1880500
QY 121 GGTGCGGAGGTCTCGCGCGGAGCTGGCGCTGTGCGGGGCGGAGTGGCGATTTGCG 180
Db 1880501 AGGGCGGAGAACTGCGCGCTGTTGCGGAGGGCGGGGCCAGGTGGCGGTCTCC 1880560
QY 181 TGTGATGTGGGAGCGCGCGGAGGTGTCGGCTGCTGAGGGGTGTTCTCTCGCGGTGT 240
Db 1880561 TGTGATGTGGCTGATCTGATGCTGCGCTGGCGGCAATGTTGGCGGGTCTGATCCGCGCTAT 1880620
QY 241 CCCTGACGCGGTGTCTGATGCGCGGTGGTGTGTCGAGCGATGCGACGATCGCCTCTC 300
Db 1880621 CCCTTAAAGGGGTGTTTCATGCGCTGGGGTGTGAGCGATGCGCGTATCAGGCGT 1880680
QY 301 ACGCCGAGCGGTGGGACGCGGTGTCGGCGGCAAGTGTGATGCGCGCTCTTTTGTGGAT 360
Db 1880681 ACACCGGATCGGTGGATACGCGGTGTCGGGCGCAAGTGTGATGGGCGCTTGGATCTGCAC 1880740
QY 361 GAGCTGACGCGGGTATGAGCTGTGCGGTTCGTGCTTCTCTCGCGCGGGATC 420
Db 1880741 GAGCTAACCGGAGCAATGGAATTTGTCGGCGTTCGTGCTTCTCTCGATGCGCGGAT 1880800
QY 421 CTGGGGTGGCGCGGAGGCAACTACCGCGGCGCAATGCGCTCTGAGCGCGCTGGCG 480
Db 1880801 GTGGGCAACCGGCTCAGGGGAATATGTCGCGCGGAATGCGTTTTCGACGGGTGCTG 1880860
QY 481 TACCGCGCGCGCGCGGTCTGCGCGGGGTGTCGCTGGGTGGGGCTGTGG 534
Db 1880861 GCCTATCGCGCTCGCGTGGGCTGGCGGATTTGTCGGTGGCGTGGGACTGTGG 1880914

RESULT 11
```

APPLICANT: Burgett, Stanley G.
 APPLICANT: Kuhstoss, Stuart A.
 APPLICANT: Rao, Nagaraja R.
 APPLICANT: Richardson, Mark A.
 APPLICANT: Rosteck, Paul R., Jr.
 TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PAUL R. CANTRELL 1138
 STREET: LILLY CORPORATE CENTER
 CITY: INDIANAPOLIS
 STATE: IN
 COUNTRY: USA
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Macintosh
 OPERATING SYSTEM: Macintosh 7.0
 SOFTWARE: Microsoft Word 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/804,198
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CANTRELL, PAUL R.
 REGISTRATION NUMBER: 36,470
 REFERENCE/DOCKET NUMBER: P9113
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-3885
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4437 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 350..14002
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 14046..20036
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 20110..31284
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 31329..36071
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36155..41830
 US-08-804-198-1

Query Match 43.1%; Score 230; DB 2; Length 44377;

Best Local Similarity 64.4%; Pred. No. 2e-35;

Matches 344; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTACGGGTGGGACGGGTGCTGGGTGGCGCGGTGGCGGATCTG 60
 Db 18618 GGACCGTGTGATACCGCGCGCACCGAGACACTGGCCCGGAAACCGCCGACCTG 18677
 QY 61 GCTGTGTGTGTGGGTGGGATCTGTGTGGTGGAGCCGGTGGTCCGGATGCTCCG 120
 Db 18678 GTGACCGGACANAGTGGCCATCTCTGTGGTGGGAGGCGGTGCCAGCACCC 18737
 QY 121 GTGTGGAGGGTGTGGCGCGGAGCTGGCGGCTGGCGGCGGAGTGGCGATTGTGCG 180
 Db 18738 GGCGTCGATCGATGCTGCGCGAGTTGACCGAGTCCGAGTCCGAGTCCCGTACGGGC 18797
 QY 181 TGTGATGTGGGGAGCGCGGAGGTGTCGGCTGCTGGAGGTGTTCTCCCGGTGT 240
 Db 18798 TGTGAGTCAAGGACCGGACCGCCCTGCGCCGCTGCTCGACGCACTCCCGGACAC 18857

QY 241 CCGCTGACGGGTGTGCTGCATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 300
 Db 18858 CCGCTGACCTGT 18917
 QY 301 AGCCCGAGCGGTGGGACGGGTGTGGCGGCGGAGGTGGATGCGGCTCTTTTGTGGAT 360
 Db 18918 ACGGCGAGCGGATCGACACGGTGTCTCGGCGCCCAAGCGCGACGCGCTCCACCTGGAC 18977
 QY 361 GAGCTGACCGCGGTGTGAGCTGTGCGGTGTGCTGTGTGTGTGTGTGTGTGTGTGT 420
 Db 18978 GAGCTGACCGCGGAGATCGAGCGGTGTGCTGTGTGTGTGTGTGTGTGTGTGT 19037
 QY 421 CTGGGTGCGCGCGGCGAGGCAACTACGCGCGGCGCAATGCGCTCTGGAAGCGGTGGCG 480
 Db 19038 CTGGGCGCGCGGCGAGGCGGCTACGCGCGGCGCAACGCTTTCATGACGCGCTGGCC 19097
 QY 481 TACCGCGCGCGGCGGCGGTCTGCGGCGGCTGTGCTGCGGTGGGGGTGTGG 534
 Db 19098 GCCCGCGGTGCGCGCGGCGGCGGCGGCGGCGGCGGCTGTGCTGCGGTGGGGGTGTGG 19151
 RESULT 8
 US-09-144-085-3
 ; Sequence 3, Application US/09144085
 ; Patent No. 6280999
 ; GENERAL INFORMATION:
 ; APPLICANT: Gustafsson, Claes
 ; APPLICANT: Betlach, Mary C.
 ; APPLICANT: Ashley, Gary
 ; APPLICANT: Julien, Bryan
 ; APPLICANT: Ziermann, Rainer
 ; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
 ; FILE REFERENCE: 30062-20020.20
 ; CURRENT APPLICATION NUMBER: US/09/144,085
 ; EARLIER FILING DATE: 1998-08-31
 ; EARLIER APPLICATION NUMBER: 09/010,809
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 33529
 ; TYPE: DNA
 ; ORGANISM: Sorangium cellulosum
 US-09-144-085-3

Query Match 41.2%; Score 220.2; DB 4; Length 33529;

Best Local Similarity 64.4%; Pred. No. 1.4e-33;

Matches 346; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

QY 1 GGGTCGGTGTGGTACGGGTGGGACGGGTGCTGGGTGGCGGCTGGCGGATCTG 60
 Db 19541 GGACCGTGTGTGATACCGCGCGGACAGGGAGCTGGGTGAGCGATCGCGCGCACCTG 19600
 QY 61 GCTGTGTGTGTGGGTGCGGATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
 Db 19601 GTGCGCGCGACGGGTGAGGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19660
 QY 121 GGTGTGGAGGTGTGTCGGGCGGAGCTGCCCGGCTGGGGGGGCGGA---GGTGTGGATTGT 177
 Db 19661 GGGGCGCGGAGTCTGTGAGTCACTCGAGGAGTCTGCGCGCGGAGACGCTGACGCTGGCT 19720
 QY 178 GGTGTGTGTGTGGGAGCGGCGGAGGTGTGCGGCTGTGCGGCTGTGTGTGTGTGTGTGT 237
 Db 19721 GCGTGTGTGTGTGAGCGGAGGTGTGCGGCTGTGCGGCTGTGCGGCTGTGCGGCTGT 19780
 QY 238 TGTCCGCTGTGAGGGGTGTGTCATGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 297
 Db 19781 GGT 19840
 QY 298 CTCAGCCGCGGCGCTGGGACGCGTGTGCGGCGCAAGGTGTGTGTGTGTGTGTGTGTGT 357
 Db 19841 CAGACGCGGCGGCTGTGCGGCGGCTGTGCGGCGGAGGTGTGTGTGTGTGTGTGTGTGT 19900

US-09-370-700-1

Query Match 44.9%; Score 239.6; DB 4; Length 80161;
Best Local Similarity 65.5%; Pred. No. 3.2e-37;
Matches 350; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTGACGGGTGGGACGGGTGTCTCGGTGCGCGGTGGCGGCATCTG 60
DB 74599 GGCACCGTGTGTACCGCGGAAACCGCGGCTCGGGCCCTGTCTCGACGCCATCTG 74658

QY 61 GCTGTGTGTGTGGGTGCGGATCTGTCTTGTGAGCGCGGTGGTCGAGTCTCG 120
DB 74659 GTACGCCCCACCGTGTGCGGACCTGTCTCTCGAAGTCTGCGGTCTGCGCGGCC 74718

QY 121 GGTGCGGAGGTCTGCGGCGGAGCTGCGCGCTTGGGCGGAGGTGCGGATGTTGCG 180
DB 74719 GGAGCGGATGAGTGTGGCGGAGCTGGAGCGCGCGCGCGCTGCGGTGTCGCG 74778

QY 181 TGTGATGTGGGAGCGGCGGAGGTGTCTCGGTCTGTGAGGCTGTCTCTCGCGGTGT 240
DB 74779 TGGACTCGGCAGATCGGACTCGCTTGGCGGCTGTGCGGTGCTGCGGAAAC 74838

QY 241 CGCTGACGGGTGTCTGATGCGGCTGTGTCTGAGCGGTGCTGAGATCGGCTCTCTC 300
DB 74839 CCGTTCGGGTGTGTGTGACCGCGCGGTGTGTCTGATGACGTGTGTCTGATG 74898

QY 301 ACGCCCGAGCGGTGCGGCGGAGTTCGCGGCGGAGGTGCGGCTCTTTTGTCTGAT 360
DB 74899 TCGCCGAGCGCTTGGACCGGTGTGCGGCGGAGGTGCGGCGGAGGTGCGGCT 74958

QY 361 GAGCTGACCGGGTATGAGCTGTGCGGTCTGTCTCTCTCGCGCGCGGATC 420
DB 74959 GAGCTGACTCGGAACTCGGTCTGTGCGGTGTGTCTCTCTCGCGCGGCTG 75018

QY 421 CTGGGTGCGCGCGGCGGCAACTAGCGCGCGGCAATGCGGCTCTGACGCGCTGCGG 480
DB 75019 TTGCGGCGGTGCGGCGGAGCAATAGCTGCGCGGCAATGCTTCTGATGCTTGGCG 75078

QY 481 TACCGCGCGCGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCTGTTGG 534
DB 75079 CATTCGCGGAGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGTTGG 75132

RESULT 6

US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991

GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

Query Match 43.1%; Score 230; DB 2; Length 44377;

Best Local Similarity 64.4%; Pred. No. 2e-35;
Matches 344; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTGACGGGTGGGACGGGTGTCTCGGTGCGCGGTGGCGGCATCTG 60
DB 18618 GGCACCGTGTGATCACCGCGGCAACCGGACATCTGGCGCGGAAACCGCGGCACCTG 18677

QY 61 GCTGTGTGTGTGGGTGCGGATCTGTCTGTGTGAGCGCGGTGGTCCGATCTCG 120
DB 18678 GTCACCGCGCACAAAGTGTGCGCATCTCTGTGTGTGGGACGCGCGGTCTCCGACGACCC 18737

QY 121 GGTGCGGAGGTCTGCGGCGGAGTGTGCGGTGTGGGCGGAGGTGGGATGTTGCG 180
DB 18738 GCGCTCGATCGACTGTGTGCGGAGTTGACCGAGTGTGCGGAGTGTGCGGCTGCGGCGC 18797

QY 181 TGTGATGTGGGAGCGCGGAGGTGTGTGCGGTGTGAGGAGGTGTCTCTCGCGGTGT 240
DB 18798 TGTGAGTACGAGCGGACCGCGCTGTGCGCGCTGTCTGACGCACTCCCGACGAAAC 18857

QY 241 CCGCTGACGGGTGTCTGATGCGGTGTGTGTGAGCGATGCGACGATGCGCTCTCTC 300
DB 18858 CCGCTGACTGTGTGTGACACCGCGGGGTGTCTGACGACGCGGTGTCTCTCGGCCAG 18917

QY 301 AGCGCGGAGCGGTGGGACCGGTGTGCGGCGGAGGTGGGATGCGGCTCTTTTGTGAT 360
DB 18918 ACGGCGGAGCGGATGACACCGGTGTCTCGGCGGCAAGGCGCGCTCCACCTGGAC 18977

QY 361 GAGCTGACCGGGGTATGAGCTGTGCGGTGTGTGTGCTGTCTCTCGCGCGCGGATC 420
DB 18978 GAGCTGACCGGGAGATGAGCGGCTGTGCGGTGTCTCTCTGTCTCTCTGCGGCCAC 19037

QY 421 TTGGGCTGCGCGGCGGAGGCAATACCGCGGCGCAATGCGCTCTTGACGCGCTGGCG 480
DB 19038 CTGGGCGAGCGGCGGCGGCGGCTACCGCGGCGCAACGCGCTTACGACGCGCTGGCC 19097

QY 481 TACCGCGCGGCGGCGGCGGCTGTGCGGCGGCTGTGCGGTGGGCGCTGTTGG 534
DB 19098 GCGCGCGGTGCGCGCGGCGGCGGCTGTGCGCTGTGCGGTGGGCGCTGTTGG 19151

RESULT 7

US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:


```
/ NAME/KEY: CDS
/ LOCATION: 31232..36067
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 36249..41774
US-08-820-227C-1

Query Match 47.0%; Score 250.8; DB 2; Length 43280;
Best Local Similarity 66.9%; Pred. No. 2.6e-39;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 GGTTCGGTGTGACGGTGGGACGGTGTGCTGGGTGCGCGGATCTG 60
DB 2295857 GGCACGGTGTGATCACCCTGGGATGCGGGTTCGGCGGTGCGGATGCTG 2295798
QY 61 GCTGGTGTGCTGGGTGCGGATCTGCTGTGTGTGAGCCGCGGTGCTCCGATGCTCCG 120
DB 2295797 GTGGCTCGTATGCGGTGCGCAATCTGGTGTGTGTGAGCCGCGGATGCTCCG 2295738
QY 121 GGTGGGAGGTCTGCGGCGCGGAGCTGCGCGGTGCGGAGGTGCGGATGCTGCG 180
DB 2295737 GGGGCTGCGGAGCTGGTGGCGGAGTGGCGGCGCGGAGTGGGATGCTGCG 2295678
QY 181 TGTGATGTGGGAGCGCGGAGGTGTGCTGGGTGCTGGAGGTGTTCTTCCCGGCTGT 240
DB 2295677 TGTGAGCGCGGATCGAGCGCGGTGGCCAAAGTGTGCGGATGTTCCGGTTCAGAT 2295618
QY 241 CCGTGTGACGGGTGTGTCATGCGGCTGTGTGTGTGAGCGATGCGATGCGCTCTCTC 300
DB 2295617 CCATTGTGCGGCTGATCCACACCGCGCGCACTCGACGACGCGGTGTGTGATGTCACTG 2295558
QY 301 AGCGCCGAGCGGTGGGACCGGTGTGCGGCGGAGGTGCGGCGGAGTGGCGCTCTTTTGTGAT 360
DB 2295557 ACACCGGATCGGTGTGATGTGTGTGCGGTGCGGCGGAGTGGCGCGGAGTGGCGCTG 2295498
QY 361 GAGCTGACCGGGGTATGAGCTGTGCGGCGGTGTGCTGTGCTGCTCGCGCGCGGATC 420
DB 2295497 GAGTTGATCGGACCTGATGTGTGCGGCTGTGTGATGTTTGTGATGCGCGGCTG 2295438
QY 421 CTGGGCTGCGCGGCGGCAACTACCGCGGCAATGCGGCTCTGAGCGCGTGGCG 480
DB 2295437 GTCCGATGTGCGGCGGCGCAACTATCGCGCGCAATTCGTTTGTGATGCGCTGGCG 2295378
QY 481 TACCGCGCGGCGGCGGCTGCGCGGCGGTGCTGCGGCGGTGCTGCGGCGGTGCGG 534
DB 2295377 GCCACCGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGTGCTGCGG 2295324

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 46.7%; Score 249.2; DB 4; Length 4403765;
Best Local Similarity 66.7%; Pred. No. 5.1e-39;
Matches 356; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1 GGTTCGGTGTGACGGTGGGACGGTGTGCTGGGTGCGCGGATCTG 60
DB 2298157 GGCACGGTGTGATCACCCTGGGATGCGGGTTCGGCGGTGCGGATGCTG 2298098

/ NAME/KEY: CDS
/ LOCATION: 31232..36067
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 36249..41774
US-08-820-227C-1

Query Match 47.0%; Score 250.8; DB 2; Length 43280;
Best Local Similarity 66.9%; Pred. No. 2.6e-39;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 GGTTCGGTGTGACGGTGGGACGGTGTGCTGGGTGCGCGGATCTG 60
DB 12864 GGCACGGTCTGCTGACGGCGCGACCGGACGCTCGGCGGCTGTGCGCGGATCTG 12923
QY 61 GCTGGTGTGCTGGGTGCGGATCTGCTGTGTGTGAGCCGCGGTGCTCCGATGCTCCG 120
DB 12924 GCGCCGAGACGGTGTGCGCATCTGCTGTGAGCGCGGCGGCGGATGCTCCGAA 12983
QY 121 GGTGCGAGGGTCTGCGGCGGAGCTGCGCGGTGCGGCGGAGGTGCGGATGTTGCGG 180
DB 12984 GCGCGGACGAACTCGCGGCGGAACTGCGCGGTGCGGAGCGGCGGTGCTTCCGCGG 13043
QY 181 TGTGATGTGGGAGCGCGGAGGTGTGCTGGGTGCTGAGAGGTGTTCTGCGCGGTGT 240
DB 13044 TGTGACCGCGGACCGGAGCGCTGCGACGCGGTGCTGCGGAGGTGCGGCGGAGCGG 13103
QY 241 CCGTGTGACGGGTGTGTCATGCGGCTGTGTGTGTGAGCGATGCGAGTGGCGCTCTCTC 300
DB 13104 CCGTGTGACGGGTGTGTCATGCGGCTGTGTGTGTGAGCGATGCGAGTGGCGCTCTCTC 13163
QY 301 AGCGCCGAGCGGTGCGGCGGAGGTGTGCGGCGGCAAGTGGATGCGGCTCTTTTGTGAT 360
DB 13164 ACCCGGACGATCGTACCGTGTGCGGCGGCGGAGCGGCGGTGTAACCTGCACTG 13223
QY 361 GAGCTGACCGGGTGTGAGCTGTGCGGCTGTGCTGTGCTGCTGCTGCGCGGCGGATC 420
DB 13224 GAATGACCGGACCGGCGGCTGTGCGGTGTTGCGGCTCTTCTGCGGCGGTGCGGCGATC 13283
QY 421 CTGGGCTGCGCGGCGGCAACTACCGCGGCGGCAATGCGGCTCTGAGCGGCTGGCG 480
DB 13284 CTGGGCGGCGGCGGCAACTACCGCGGCGGCAACTACCGCGGCGGCAACTCTTCTGAGCGGCTGCGG 13343
QY 481 TACCGCGCGGCGGCGGCTGCGCGGCGGTGCTGCGGCGGTGCTGCGGCGGTGCGG 534
DB 13344 CAGCACCGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGTGCGGCGGTGCGG 13397

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 47.0%; Score 250.8; DB 4; Length 4411529;
Best Local Similarity 66.9%; Pred. No. 2.5e-39;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 23.1323 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggctggtgtggtgacggg.....cgtggtggtggggtgtggtg 534

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/FCUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	250.8	47.0	43280	2	US-08-804-227C-1
2	250.8	47.0	4411529	4	US-09-103-840A-1
3	249.2	46.7	4403765	4	US-09-103-840A-2
4	239.6	44.9	80161	3	US-09-036-987A-1
5	239.6	44.9	80161	4	US-09-170-700-1
6	230	43.1	44377	2	US-08-804-227C-7
7	230	43.1	44377	2	US-08-804-198-1
8	220.2	41.2	33529	4	US-09-144-085-3
9	218.8	41.0	4403765	4	US-09-103-840A-2
10	218.8	41.0	4411529	4	US-09-103-840A-1
11	212.4	39.8	15872	4	US-09-105-537-1
12	199.6	37.4	13842	4	US-09-105-537-30
13	199.6	37.4	36778	4	US-09-105-537-5
14	199.6	37.4	38506	3	US-09-320-878-19
15	198	37.1	13987	2	US-08-804-227C-13
16	196.4	36.8	11220	4	US-09-105-537-32
17	192	36.0	20235	1	US-07-642-734C-3
18	192	36.0	20235	3	US-08-439-009A-3
19	183.6	34.4	68750	3	US-09-335-409-1
20	183.6	34.4	68750	4	US-09-568-102-1
21	183.6	34.4	68750	4	US-09-567-969-1
22	183.6	34.4	68750	4	US-09-568-480-1
23	183.6	34.4	68750	4	US-09-568-486-1
24	183.6	34.4	68750	4	US-09-568-472-1
25	183.6	34.4	68750	4	US-09-567-899-1
26	182	34.1	71989	4	US-09-443-501A-2
27	178.8	33.5	28958	1	US-08-258-261B-6

28	178.8	33.5	28958	1	US-08-456-837-6
29	178.8	33.5	28958	1	US-08-457-342-6
30	178.8	33.5	28958	1	US-08-457-646A-6
31	178.8	33.5	28958	1	US-08-458-076A-6
32	178.8	33.5	28958	1	US-08-764-233A-4
33	178.8	33.5	28958	1	US-08-457-335A-6
34	178.8	33.5	28958	1	US-08-729-214-6
35	178.8	33.5	28958	3	US-09-028-934-6
36	178.8	33.5	49377	1	US-08-764-233A-1
37	174	32.6	11219	1	US-07-642-734C-1
38	174	32.6	11219	3	US-08-439-009A-1
39	166	31.1	50937	4	US-09-428-517-1
40	158.8	29.7	477	3	US-09-010-809-2
41	158.8	29.7	477	3	US-09-010-809-17
42	132	24.7	4689	4	US-09-105-537-34
43	90.4	16.9	897	4	US-09-434-288-6
44	81.4	15.2	1681	4	US-09-434-288-7
45	72	13.5	320	4	US-09-165-264-13

ALIGNMENTS

RESULT 1
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:

JOURNAL

Patent: WO 0001827-A 1 13-JAN-2000;
KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
STAUTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
HAMISH ALASTAIR IRLIN (US); STUTZMAN ENGMALL KIM JONELLE (US)

FEATURES

source

1. 12381
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

BASE COUNT 1884 a 4561 c 4005 g 1931 t

ORIGIN

Query Match 100.0%; Score 534; DB 6; Length 12381;
Best Local Similarity 100.0%; Pred. No. 5.1e-69;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTTGTAACGGGTGGAGCGGGTGTGCTGGTTCGGCGCGGCGGCAATCTG 60
DB 11426 GGGTCGGTGTGTTGTAACGGGTGGAGCGGGTGTGCTGGTTCGGCGCGGCGGCAATCTG 11485
QY 61 GCTGTGTGTGTGGGGTGCAGGATCTGCTTTGGTGTAGCCGGCTGGTCCGAGATCTCCG 120
DB 11486 GCTGTGTGTGTGGGGTGCAGGATCTGCTTTGGTGTAGCCGGCTGGTCCGAGATCTCCG 11545
QY 121 GGTGCGAGGCTGTGCGGGCGAGCTGGCCGCTTGGGGCGGAGGTGCGAATTGTTCG 180
DB 11546 GGTGCGAGGCTGTGCGGGCGAGCTGGCCGCTTGGGGCGGAGGTGCGAATTGTTCG 11605
QY 181 TGTGATGTGGGGAGAGCGCGGAGGTGTGTCGGCTGTGAGAGGTGTCTTCCGCGGGTGT 240
DB 11606 TGTGATGTGGGGAGAGCGCGGAGGTGTGTCGGCTGTGAGAGGTGTCTTCCGCGGGTGT 11665
QY 241 CCGCTACAGGATCTGCTGATCGGCTGTGTGTCTGAGACGATCGACATCGCTCTCTC 300
DB 11666 CCGCTACAGGATCTGCTGATCGGCTGTGTGTCTGAGACGATCGACATCGCTCTCTC 11725
QY 301 AGCCCGAGCGGCTGGGCAACGATGTCGCGGCCCAAGGTGATCCGCTTTTGTCTGAT 360
DB 11726 AGCCCGAGCGGCTGGGCAACGATGTCGCGGCCCAAGGTGATCCGCTTTTGTCTGAT 11785
QY 361 GAGCTACCGGGGTATGAGAGCTGTGCGGCTTGTCTCTCCGCGCGGCGGAGATC 420
DB 11786 GAGCTACCGGGGTATGAGAGCTGTGCGGCTTGTCTCTCCGCGCGGCGGAGATC 11845
QY 421 CTGGGCTCGGCGCGGAGCACTACGCGCGGCAATGCCCTCTGAGACGCGCTGGCG 480
DB 11846 CTGGGCTCGGCGCGGAGCACTACGCGCGGCAATGCCCTCTGAGACGCGCTGGCG 11905
QY 481 TACCGGCGGCGGCGGCTGTGCGGGGCTGTGCTGCGCTGGCGCTGGCGCTGGCG 534
DB 11906 TACCGGCGGCGGCGGCTGTGCGGGGCTGTGCTGCGCTGGCGCTGGCGCTGGCG 11959

RESULT 2

E38020 30690 bp DNA linear PAT 31-JAN-2002

LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909

KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE

AUTHORS Omura, S. and Ikeda, H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;
THE KITASATO INSTITUTE

COMMENT

OS Streptomyces avermitilis
PD JP 2000245457-A/1
PR 12-SEP-2000
PR 24-FEB-1999 JP 1999046961

PI SATOSHI OMURA, HARUO IKEDA
PC C12N15/00, A61K31/70, C12N1/15, C12N9/88, C12P19/62, PC
C12Q1/68/C07D493/22.
PC (C12N1/15, C12R1:465), (C12N9/88, C12R1:465), (C12N15/00 CC
FT Key Location/Qualifiers
FT CDS (1), (11916)
FT CDS (11971), (30687).

FEATURES

source

1. 30690
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

BASE COUNT 5355 a 12455 c 8617 g 4263 t

ORIGIN

Query Match 100.0%; Score 534; DB 6; Length 30690;
Best Local Similarity 100.0%; Pred. No. 4.3e-69;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGTTGTAACGGGTGGAGCGGGTGTGCTGGTTCGGCGCGGCGGCAATCTG 60
DB 10609 GGGTCGGTGTGTTGTAACGGGTGGAGCGGGTGTGCTGGTTCGGCGCGGCGGCAATCTG 10668
QY 61 GCTGTGTGTGTGGGGTGCAGGATCTGCTTTGGTGTAGCCGGCTGGTCCGAGATCTCCG 120
DB 10669 GCTGTGTGTGTGGGGTGCAGGATCTGCTTTGGTGTAGCCGGCTGGTCCGAGATCTCCG 10728
QY 121 GGTGCGAGGCTGTGCGGGCGAGCTGGCCGCTTGGGGCGGAGGTGCGAATTGTTCG 180
DB 10729 GGTGCGAGGCTGTGCGGGCGAGCTGGCCGCTTGGGGCGGAGGTGCGAATTGTTCG 10788
QY 181 TGTGATGTGGGGAGAGCGCGGAGGTGTGTCGGCTGTGAGAGGTGTCTTCCGCGGGTGT 240
DB 10789 TGTGATGTGGGGAGAGCGCGGAGGTGTGTCGGCTGTGAGAGGTGTCTTCCGCGGGTGT 10848
QY 241 CCGCTACAGGATCTGCTGATCGGCTGTGTGTCTGAGACGATCGACATCGCTCTCTC 300
DB 10849 CCGCTACAGGATCTGCTGATCGGCTGTGTGTCTGAGACGATCGACATCGCTCTCTC 10908
QY 301 AGCCCGAGCGGCTGGGCAACGATGTCGCGGCCCAAGGTGATCCGCTTTTGTCTGAT 360
DB 10909 AGCCCGAGCGGCTGGGCAACGATGTCGCGGCCCAAGGTGATCCGCTTTTGTCTGAT 10968
QY 361 GAGCTACCGGGGTATGAGAGCTGTGCGGCTTGTCTCTCCGCGCGGCGGAGATC 420
DB 10969 GAGCTACCGGGGTATGAGAGCTGTGCGGCTTGTCTCTCCGCGCGGCGGAGATC 11028
QY 421 CTGGGCTCGGCGCGGAGCACTACGCGCGGCAATGCCCTCTGAGACGCGCTGGCG 480
DB 11029 CTGGGCTCGGCGCGGAGCACTACGCGCGGCAATGCCCTCTGAGACGCGCTGGCG 11088
QY 481 TACCGGCGGCGGCGGCTGTGCGGGGCTGTGCTGCGCTGGCGCTGGCGCTGGCG 534
DB 11089 TACCGGCGGCGGCGGCTGTGCGGGGCTGTGCTGCGCTGGCGCTGGCGCTGGCG 11142

RESULT 3

AB032367 64957 bp DNA linear BCT 18-SEP-1999

LOCUS AB032367 64957 bp DNA linear BCT 18-SEP-1999
DEFINITION Streptomyces avermitilis polyketide synthase gene cluster (aveA1, aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION AB032367
VERSION AB032367.1 GI:5902890

KEYWORDS AveC, cytochrome P450 hydroxylase; type I polyketide synthase AVE5
4; type I polyketide synthase AVE3; type I polyketide synthase
AVE5 2; type I polyketide synthase AVE5 1.
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE

AUTHORS Ikeda, H., Nonomura, T., Usami, M., Ohta, T. and Omura, S.
TITLE Organization of the biosynthetic gene cluster for the polyketide
anthelmintic macroide avermectin in Streptomyces avermitilis

QY 421 CTGGGGTCCGCGCGGCACTACGCCCGCCCAATGCCCTCTGACGCGCTGGCG 480
 DB 19705 CTGGGGTCCGCGCGGCACTACGCCCGCCCAATGCCCTCTGACGCGCTGGCG 19764
 QY 481 TACCGCGCGCGCGCGCGCTCTGCCGGGGGTCTCGTGGCGTGGCGCTGG 534
 DB 19765 TACCGCGCGCGCGCGCGCTCTGCCGGGGGTCTCGTGGCGTGGCGCTGG 19818

RESULT 5
 AB032367/c 64957 bp DNA linear BCT 18-SEP-1999
 LOCUS Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
 aveA2, aveA3, aveA4) and aveC, aveS genes, complete cds.
 ACCESSION AB032367
 VERSION AB032367.1 GI:5902890
 KEYWORDS aveC; cytochrome P450 hydroxylase; type I polyketide synthase AVEs
 4; type I polyketide synthase AVEs 3; type I polyketide synthase
 AVEs 2; type I polyketide synthase AVEs 1.
 SOURCE Streptomyces avermitilis DNA.
 ORGANISM Streptomyces avermitilis
 Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;
 Streptomyces; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 1 to 64957)
 Ikeda, H., Nonomura, T., Ueami, M., Ohta, T. and Omura, S.
 Organization of the biosynthetic gene cluster for the polyketide
 anthelmintic macroide avermectin in Streptomyces avermitilis
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
 99380548
 JOURNAL 2 (bases 1 to 64957)
 MEDLINE Ikeda, H., Nonomura, T., Ueami, M., Ohta, T. and Omura, S.
 DIRECT SUBMISSION
 AUTHORS Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
 JOURNAL Sciences, Kitasato University, Microbial Chemistry; 5-9-1
 Shitokane, Minato-ku, Tokyo 108-8641, Japan
 (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242,
 Fax: +81-3-3444-6197)
 FEATURES
 source Location/Qualifiers
 1. 64957
 /organism="Streptomyces avermitilis"
 /db_xref="taxon:33903"
 101..12019
 /gene="aveA1"
 101..12019
 /gene="aveA1"
 /note="contains a loading module, Ave module 1 and module
 2
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVEs 1"
 /protein_id="BA84474.1"
 /db_xref="GI:5902891"
 gene
 CDS
 101..12019
 /gene="aveA1"
 /note="contains a loading module, Ave module 1 and module
 2
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVEs 1"
 /protein_id="BA84474.1"
 /db_xref="GI:5902891"

VMSVVDILRRDAGDAWBERADVQVLFVWVSLAALRSYGIIEPDVILGSHSGEIA
 AAVHGCALSILKOAATKVALASRALAVRRGASVAPLAPOEVBOLIGRRMGRLLVA
 AVNGPRSTAVSGDAEAVDEVLAVCAGTGRRARIPDVASHCHPNOLBELLELGD
 ISPOBSGVPESTVEGTWIDTTLDAAYVYRNLHQVRSDVAQALADGHRVFEVS
 PHPTLVPLIEDTTEDEAVETAIIGSLRRDNDOTRRLLTALHTHTTGTPPTWHAHY
 THNHTPHHTLIDLETYPFOHQWLESQBSQBSGAGSGAGSGAGSGAGSGAGSGAGTA
 EVESRFWDVAVARODLETVAATTLAVPSALDVTVPALSMHNOQOALINTVGET
 MKPLTLPTTHORHOTWILAIPEQONHPHITLITLTHNHGHTLPILTLNHTYNOH
 LHTHTLHTHQOAOHNTTGAITGLLSILADETPHPHHTPTGTLNLITLTONPOT
 PPTPLVATTNATLTHNDPLTHPTQAOHWGLARTTLLEHPTHTAGIIDLPTPTPT
 LOHLTOQLTPHNOTOLAIRTTGTTRRLPTTLTPHPTQPTPTGTLITGTETAL
 ATHTLHTLTHPTQHTLITSTRGTPTPAQHLLTQLOQGHILITLITDSDPDLO
 QMLNTIPQHPHTVITHTAGILIDATLITLTLTQNLNVRAKASHNLHQLPOTPL
 TEVLVYSSAAATFGAGQANVAAVYDALAHNHTHTLITLITSMGTMSGGLDLS
 DKARVILDRGPRMSPELATVAVQALIDTERPVHLLIDMSKTENHSQSDLYSA
 ABERBAVORPTPRALHTLTAHOTBSADRALLEVRHVAVALRHADPKAIAPQDS
 FRALGDSLTAVEFRNLTKATGLIPSLVDHPPTPALAVLONQLGTAESAPS
 AAVALAEASVTEPRIAIVGMACTPRGGLVSDPFMDLISEODALIGEPTRDGLDITL
 YPDDHDEGTCTYRNGFLYDAGHDAEFGLSPREBALMDPOQLLLETAMETIEHA
 GINPHTLHOTPTGTFGTGODYALRVNAGSGSTGFAALGTAGVLSGRISYTPGE
 GPVAVDTRACSSILVALHIALCOLARAGCSMALAGVTWMSRGAVERSRORGLAD
 CHXKAPSAADDTGEGVGMLVERLSDAHNGRVLAVGSAVNOGASNGLTAP
 NPSQOQVLRQALANAGSAGVDVAEAGTGTTLDDPLLEAOALATYGODRAGEPL
 MGSVKSNGVTRQAAAGVAGVIMVVALHNGLLPTLHVDEBSPHVMSAGAVQLLLE
 TVPMEGEGRLRRAGVSPFGVTNAHVLEBAPADVGPAGBAGBAGDEBAAG
 SGVMPVLVSAKSOPALRAOQALAHNLDHGLDLDADVGYTLAHARVAFDRATLIA
 ADPRFLOALQALAGEPRVAVIHSSAREGTGCAAGTATISCGQGTORPMAGL
 YHTHFVFAALNDITCTHIDPHLPLRLTQNDNDENRAALLQOTRYAOPLRFQ
 VALHRLTDGNIITHTYVAGSLGITAHLNLGLITLDTATLITORATLMOTPEGT
 MTLHTHTPHITHTLTAHENDLAIAGITLVISGTPTVQHTLITLCOQOQIKKT
 LPTNHAFSHPNPIILNLOLHOTLTLYHPPTPTITANTPDOLLTPHMYQOANT
 VYATTTOTLHOHGVTYIELEBDTLTLTHNHLNPTTLTLTHPHNHTOHTLT
 NIAKTTTMMPHHTHNDONPHTHTHLDIPYPROHNHYLESSTOPGANNVAAQIDP
 TENPILGATLEATTGCGALNRLSLRSHRPLADHNAVGTVLVLSGTLLELDHAGTY
 VECDDVDELITLAPLVVPEVDSVQVQVAAADGGRRLVSVYARGGASGCGGASAG
 VWTCHASGLVLEAAAGVAVDLAGVPRGAVAAVDVDRPAGMAGAGVCLPVEFGL
 RAVMRDGDILAEVCLPEBAMGDAGFGILHPLAGVOPRLVPLPGCFPESGAFG
 EGVRAVWAGVSLRAGVTGVRVAVSAGRGSGAVSVVVDGSGAVVAVSDRLT
 RPYMDGOLRAVSVSARGSLYAVOMAVGVPVPGVGOAMAMHEDUCEGSGGVV
 LRCPRDAGCGCGCGCGGCGVCGVCGVCGVCGVGLGELERFASRLVYVTRKAVAPPE
 DGPVVVAGVWGLVRSQAEHPRDFVLDLDTDTGTDLTAGAGMAGVDRGVAHV
 AGCEPOLVWGERLLAARLKLESSGDVPAODSGTTRARSVPAPORSOGVAVRVD
 VSGREVLPMISGSGVLTGCTGVLGAALVHLAGVCGVVDLTVLSRGGDAPCAEGLR
 AELALGAEVRIVACDVGREVRVLLSGVPGCLTVGVNAAVGLDADTITSLPER
 ICTVRAKVYDALLLDELTRGHELSAPILFSAACILISAGCGNVAANAALDALAYR
 RRAAGYPGVSLAMGLEBSAGHTGLAGDHRKILIRSGLHPTSDPALLPDAALRQ
 REVLLPADLRAPAPLPEPLLODLIPATRRRTTITTTGADNCAOLHARLACQTHEQH
 TTLALVRSNHTVIGLHTTTPDTIPEDRAFRDGLPSLTAVEILNRSLRTTGLRLPTTL
 AFDHNPPTLTHLHTLOPOPDNAVAVPLAELDKLSALSLADTDSBSBVTTLRLX
 SLMLRWNPQHPTASADDDEKFTSATAEIKFKIINDIGLS"
 12071..30790
 /gene="aveA2"
 12071..30790
 /gene="aveA2"
 /note="contains Ave module 3, module 4, module 5 and
 module 6
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVEs 2"
 /protein_id="BA84475.1"
 /db_xref="GI:5902892"
 gene
 CDS
 12071..30790
 /gene="aveA2"
 /note="contains Ave module 3, module 4, module 5 and
 module 6
 multifunctional polyketide synthase"
 /codon_start=1
 /transl_table=11
 /product="type I polyketide synthase AVEs 2"
 /protein_id="BA84475.1"
 /db_xref="GI:5902892"

NDICTLDBLHPLPLTODPNTODTTLEBAALLOOTRAYAPALFAOVALHRL
 LDGVTHTPHYAGSHLGETTAHLAGILTLADATTLTORALOTMPGPTTLTAT
 TPHHTHNLTAHENDLAAINTPSLVSIGPHVOTHTLCOOOGITKTLPTNHA
 FHSPTNPILNOHOTOHTLTPHTPTPLTANTPDDLTTPHTYQOARNTDYAT
 TMLHOHGVTTYELGPDNTLTTLHNDLNPHTTTLTLHPHNOHTLNLATYT
 TTPHHPYHNNQPHHTHLDLPTPFOHNNWLOPCKSPSPSEBEOELQATST
 PLADYVGSPOREDEBLRLVTAHAAYGHATPREVIYENKFKELGDSLAOLR
 NRLADVDLPLATLFDYPTPMAJCOPLAAVAGDTTGTTLTPADPAPLIV
 GMACRYBGDPTDLMOVSGGHDALGPPNRMGDLOTIYNPDNDHGTSTSGG
 FLYDAGNPDPTGDISPRALANDPOORILLERTAMSIEHACINPSLSTPGVAG
 LTYHDNAAPPTAPAGFEGYLGHSAGSLASGVAVALGEGALVVDACSSVAL
 HLAQALRSCEGSMALAGVTVSTAGVEFROGKAFSAADGTGEGE
 GVMGLVERLSDBARLGRHILAVRGSAVODASGSLAPNPSERYIALANAD
 LPAVDVAEAGHTGTTGDPLEAOLALLTGODRGNGNPLGSKSNIGHAAG
 VSGVITKMMALNGLPLTLHNDERSPHYDMSAGAVOLLETYPMGEGGRILRAAS
 SSGVSTNAHVILEBAPANIPEDTPADAPGSAADVDYGEAGSADASTGEBATPA
 AASPGVFWLVSASQOPALAOALHATLDBPGLADVGTTLHAARVDRHRTL
 LAADRTFLOALALAGBHPAVIHSSAPGTGEBAGKTAFCISGQGTORPMAN
 GLYHTPVAALANDICTLDBLHPLPLTODPNTODTTLEBAALLOOTRAYA
 PALFAOVALHRLLDGVTHTPHYAGSHLGETTAHLAGILTLADATTLTORALM
 OMPBGTMTLHTPHTHNLTAHENDLAAINTPSLVSIGPHVOTHTLCOO
 OGITKTLPTNHAHSPTNPILNOHOTOHTLTPHTPTPLTANTPDDLTTPHTY
 TQOARNTDYATTTTLHOHGVTTYELGPDNTLTTLHNDLNPHTTTLTLHPH
 POTHNLNTAKTTLTPHHPYHNNQPHHTHLDLPTPFOHNNWLOPCKSPSP
 SSGSGRACGTAGTAVERFMDAVARODLETVALAVPSAGSLDTPALSAMHR
 OHDOARINTWPTOEKPLPLTPOHOMOLAIPEOTHHPHITLNLHNGIT
 PLILNHTHTNDPHTROAONHTPRTIGLILALDERPHNHHTPTGTLN
 LITOTHTOTHTPTPLMAYATNTKATTHPNDPLHTPOATOGKATTLLENHTHNGI
 IDLPTTPHTHNLTLTOLLOPHOTOLAIRTTGHTHRLTLHOOKPHLPT
 TTLITGEGALATLHNLTLTPOHTLHLSRTGHTPHAOHTLLOOKPHLPT
 TCDSTNDPOLQALNTIPEONPLTVIHAGLIDDLATLNLPTOLANLPLAKASHM
 LHLTHTPHTPLQALNTIPEONPLTVIHAGLIDDLATLNLPTOLANLPLAKASHM
 GIMOGGSLATGOVSEHLRRRGAMPBELAVTAVOLASIGRSPHILADIMKXIGV
 LSKSVLIEDLPQOAGTEBASTVQSTYRLQOLMERSSEBOEELISVRIHSA
 AYLGRDSEALPPEGLRDLGDSLAVALRNLALQOTELATPLTVEDPSPKLAQ
 FLISLAEFOQDNSTPLPRRALEDEPLIAIVGACRFPGVTSADPFWLISSEDAI
 GEPTRGMDLDTLXDPDHPGTCYTRNGEPLUDYADVAGSADPFWLISSEDAI
 RLLETAMEITIEHAGINPTLHGTPTGVTGNOOAHAIROAPSTGSGFVLGAT
 SIASGRISYILIEGPAVTLDTACSGSLAHLAHOOSLHAGTLAGAIVMTPTI
 TTEPARORGLAPDRCSKAFSAADGTGEGGMLVRLSDBARNGRVLAVRGS
 ANODASNGLTBNPSSOQVTRLOALALADPVDVAWEAGTTLTGLEADAI
 LATYQDRPENGPLMIGSVKSVNGTQOAAAGAVIKNVLAHRLPTLTLADDESP
 HDWSAGAVOLLETYPMGEGGRPRRAVSSGVSGLANHLVLEBAPADVDYGEPA
 DBDASGEBAAGSPGVMPLVSAKOPALRAOALHNLHDLHPDLADVDYGEPA
 ABAVDRHARLTLAADRTFLOALALAGBHPAVIHSSAPGTGEBAGKTAFCIS
 GQGTORPMANGLYHTPVAALNDICTLDBLHPLPLTODPNTODTTLEBAAL
 LOOTRAYAPALFAOVALHRLLDGVTHTPHYAGSHLGETTAHLAGILTLADAT
 ILTORALMOTMPGPTTLTATTPHHTHNLTAHENDLAAINTPSLVSIGPHVOT
 HITTLCOOOGITKTLPTNHAHSPTNPILNOHOTOHTLTPHTPTPLTANTPDD
 OLLTPHTYQOARNTDYATTTTLHOHGVTTYELGPDNTLTTLHNDLNPHTTTL
 TLHPHNOHTLNLATYTTPHHPYHNNQPHHTHLDLPTPFOHNNWLOPCKSPSP
 PSAOTSROGRSRSRSPDTASEFMDAVNEBLOSALFTLIDALADVDYVALSAMH
 RHOHDARINTWPTOEKPLPLTPOHOMOLAIPEOTHHPHITLNLHNGIT
 LPTPLTVHNTNPQHLHNTROAONHTPRTIGLILALDERPHNHHTPTGTLN
 TQOARNTDYATTTTLHOHGVTTYELGPDNTLTTLHNDLNPHTTTLTLHPH
 TTAGLIDPLPTPTHTLHNLTLTPOHNOALIRTTGHTHRLPTLTLPTPOH
 TTPHHPYHNNQPHHTHLDLPTPFOHNNWLOPCKSPSPSEBEOELQATST
 INHTITCTSPNDLOQLNTIPEONPLTVIHAGLIDDLATLNLPTOLANLPLAKASHM
 KATGAALHLELDBETLHFLIFSSGAGWSGNCASANAAYDALATIRTHQHL
 PASIASMPWAGKMSAGDAAGYLEKREILPEMPMALAFTTAAQOAPNSBLRIAD
 IMERVPAPFTARRSHPLIEDIPEVQAOAELEASTAKTAPRIASBLERLAL
 TSSKONOVLLIGRTGICVGLNBEGBEIDORAPDLGDSLTASQPEKELAKGL
 PLPSLIVDPPTPOBCAHLRTQVLYNDDEBAAISNALPOVARNSTQPEKELAKGL
 ACRRPGVARSADDLMEILASGKDALGVFTTDGMDLDTLYDPPHPTCYTRNGGFL
 YAGHDAEFPGISPREALAMPQOOLLETWMETIEHAGINPTLHGTPTVAGIN
 AODHAHROSBRVETIEBGVALTGSSGSAVASTLDEBPAVSDVTRACSSIVAL

Query Match 99.7%; Score 532.4; DB 1; Length 64957;
 Best Local Similarity 99.8%; Pred. No. 6.3e-69;
 Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GGGTCGGTGTGGTGA CGGGTGGACCGGTGTGCTGGTGGCGCGGTGGCGCGCATCTG 60
 |||||

Db 45573 GGGTCGGTGTGGTGA CGGGTGGACCGGTGTGCTGGGTGGCGCGGTGGCGCATCTG 45514
 QY 61 GCGTGGTGTGGTGGTGGTGGGATGCTGCTGTGTGTGAGCGCGGCGGTGTCGGATGTCGG 120
 Db 45513 GCGTGGTGTGGTGGTGGTGGGATGCTGCTGTGTGTGAGCGCGGCGGTGTCGGATGTCGG 45454
 QY 121 GGTGCGAGAGGTCGTGCGGCGGAGAGCTGCGCGGCTTGCGGCGGCGAGGTCGATTTGCG 180
 Db 45453 GGTGCGAGAGGTCGTGCGGCGGAGAGCTGCGCGGCTTGCGGCGGCGAGGTCGATTTGCG 45394
 QY 181 TGTGATGTGGGAGAGCGCGGAGAGTGTGTCGCGCTGTGTGAGAGGTCCTGCGCGGT 240
 Db 45393 TGTGATGTGGGAGAGCGCGGAGAGTGTGTCGCGCTGTGTGAGAGGTCCTGCGCGGT 45334
 QY 241 CGGTGACGGGTCGTGTGATGCGGTCGCTGTGTGAGAGAGTGCAGATGCGATGCGCTCTTC 300
 Db 45333 CGGTGACGGGTCGTGTGATGCGGTCGCTGTGTGAGAGAGTGCAGATGCGATGCGCTCTTC 45274
 QY 301 ACGCCGAGCGGCTGCGGCA CGGTCGCTGCGGCGCAAGTGTGATGCGCTCTTGTGTGAT 360
 Db 45273 ACGCCGAGCGGCTGCGGCA CGGTCGCTGCGGCGCAAGTGTGATGCGCTCTTGTGTGAT 45214
 QY 361 GAGCTGACGGGAGTATGAGAGCTGTGCGGCTGTGTGCTGTCTCTCTGCGCGCGGAGATC 420
 Db 45213 GAGCTGACGGGAGTATGAGAGCTGTGCGGCTGTGTGCTGTCTCTCTGCGCGCGGAGATC 45154
 QY 421 CTGGGGTCGGCGCGGAGGCAATACCGCGGCGCAATCGGCTGTGAGCGGCTGCGG 480
 Db 45153 CTGGGGTCGGCGGAGGCAATACCGCGGCGCAATCGGCTGTGAGCGGCTGCGG 45094
 QY 481 TACCGGCGGCGGCGGCTGCGCGGAGTGTGCGCTGCGGCGGCTGCGGCGGCTGCGG 534
 Db 45093 TACCGGCGGCGGCGGCTGCGCGGAGTGTGCGCTGCGGCGGCTGCGGCGGCTGCGG 45040

RESULT 6
 SGR300302/c 39314 bp, DNA linear BCT 20-MAR-2002
 LOCUS Streptomyces griseus partial ORF1, cana gene, canc gene, canf gene,
 DEFINITION canf gene, canra gene and canb gene.
 ACCESSION AJ300302
 VERSION AJ300302.1 GI:12231153
 KEYWORDS ABC-transposase; aminotransferase; can RB gene; cana gene; canc
 gene; canf gene; canra gene; canc gene; cho-like protein;
 cytochrome P-450; ferredoxine; PABA synthase; thioesterase.

SOURCE
 ORGANISM Streptomyces griseus.
 Streptomyces griseus.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 REFERENCE
 1 Campelo, A. B.
 Thesis (2000) Department of Microbiologia, Universidad de Leon,
 Leon, Spain
 JOURNAL
 REFERENCE
 2 Campelo, A. B. and Gil, J. A.
 The candidicin gene cluster from Streptomyces griseus IMRU 3570
 Microbiology 148 (Pt 1), 51-59 (2002)
 MEDLINE
 21642576
 11782498
 3 (baaes 1 to 39314)
 REFERENCE
 GIL, J. A.
 TITLE Direct Submission
 JOURNAL Submitted (14-NOV-2000) Gil, J. A., Microbiologia, Universidad de
 Leon, Campus de Vegazana s/n, 24071, SPAIN
 FEATURES
 source
 1. 39314
 /organism="Streptomyces griseus"
 /strain="IMRU 3570"
 /db_xref="taxon:1911"
 complement(1..1785)
 /note="ORF1"
 /codon_start=1
 /transl_table=11

CDS

1 GGGTCGGTGTGGTGA CGGGTGGACCGGTGTGCTGGTGGCGCGGTGGCGCATCTG 60
 |||||

/product="Cno-like protein"
/protein_id="CAC22112.1"
/db_xref="GI:12231154"
/db_xref="SPTREMBL:O9EWC6"
/translation="MLERTTALVDAALGATTEGSSLVLTGPVGIGRSALQRL
SLCADDVRLVRAHAAPMEDFAFVVRQULETLLGDSPEDARDMDAHAFARH
VLADDAAPPGADQALNATEAVLHGLSLANVSADRLILVDLQMSDPAVLRWLT
LAKRLHLRAVVVCAIRDGDPSSHHTLVREIRAAQTLPASLSLTAHLEVRHFG
EAGDDEFVOACHGVGNPLFLSLVGTGFGRRPLAEHAEARLRPSOLRERLAS
ILRTOPAPVRDLAAAIILGEOSDAPTLARLGLDSIGYAGALRALGALGALAPDP
RPIHRSVRDAESTLTVMORERHMDAAALLVYAGSPAQVAAQLMAVYTRQPNAYD
VRAADDTLRKGAFTAGYLCRALDSDPAGVNGRLLVELTHERGPDPLACRRA
IAQAMTLLEPRDRAVAALRISPTAGPPLTAVDLRQAAEDLGPADLGTDRDLA
LLEALRLRHCGHEDPHELAASARLKELGPEPPVDTGGERELTIALLGATLGCILPA
SEVYLAGRVLDDEPATSARISHSTLPLAI"
9525..10583
/gene="cana"
9525..10583
/gene="cana"
/codon_start=1
/transl_table=1
/product="aminotransferase"
/protein_id="CAC22113.1"
/db_xref="GI:12231155"
/db_xref="SPTREMBL:O9EWC5"
/translation="MAFTHPSRPLDGRLELYSDAVSGNISQGPYVRFEENAF
ENGVHAGVACSGTALTLALNALINIGDEYVEFTMVASAMAVTGTATPAVVD
CGDDINDVTRIEEKITPRTRAVMPHAYGRCDMAVMDLQVNLRVDESAEAG
VARPGDIACFSLPANKITRAGEGCVLTDDPRLEALHLRMAAFTDRDSFLHKILAY
NTRMTMGAVALAQTERRLEILATREIEARDALKOLPGITLMPADVLMWDLR
ASREBLRAHLDEGIETRLFPKMSRQGVDPVWPPTNAHRSFEDGLYLPHTGTL
AADQEYITGAVRDPFYAG"
10678..11859
/gene="cana"
10678..11859
/gene="cana"
/codon_start=1
/transl_table=1
/product="cytochrome P-450"
/protein_id="CAC22114.1"
/db_xref="GI:12231156"
/db_xref="SPTREMBL:O9EWC4"
/translation="MTTSPGPTVDPDPRTPREPLPLSQYAEHRKQNGLVQTHLPGNR
PIMLVTRHEDVAVLTHPRISANPDNEGPNVGETGVKQEQIPGMFVGLSPEDR
PKVLIPERTVRRVRELRAPIERTVDERIDAMLAGCTADLVDPALPVPISALIL
GVPSADRDPEESTRTLVARTSTDEERABARQLRYINRLVYKKRGEDILSRIL
LSTGKDSBELSGVLLLLIAGHETTANNIGGVVTLISHREVTGDRLVEBLRLHS
VADMVALRAVADVETIAGQTIKRGEGIVPLASANDTEAFGCPHAFNPERTERRIVA
FGYGVHCIGQNLVVRVEMEIAYRKLPERIPELRVAVPEDQLAYKYDGIILFGLHELPRV
W"
11906..12100
/gene="canf"
11906..12100
/gene="canf"
/codon_start=1
/transl_table=1
/product="ferridoxine"
/protein_id="CAC22115.1"
/db_xref="GI:12231157"
/db_xref="SPTREMBL:O9EWC3"
/translation="MRVTVDSEQCVAGOCVLNAPFVFDODDGVVLLRAPEBDRDH
EAVRTAGDLCPSASVVLQED"
12223..12993
/gene="canf"
12223..12993
/gene="canf"
/codon_start=1
/transl_table=1
/product="thioesterase"
/protein_id="CAC22116.1"
/db_xref="GI:12231158"
/db_xref="SPTREMBL:O9EWC2"
/translation="MTTADDTAARWLRRYHPAEADAVRLVCPFHAGSASFYHPVPSAR

FAPGAEVSLQYGRQDRKSCVCELDGLADLTLEQLPLDERPTVFEHSGMALA
FETARMLEQKAGPRTVIASGRGSTRRAEVRDGDGIAEKKRLNGTGAAGVGD
BEILARMLPALRGDPRALLETYCCPDRDRACTLVYTGSDPLTTVEAEKRRDHTTG
PRLRLFTGCGHFLQHDVAVTEIAQLHPRAAPAA"
13190..15361
/gene="pabab"
13190..15361
/gene="pabab"
/codon_start=1
/transl_table=1
/product="PABA synthase"
/protein_id="CAC22117.1"
/db_xref="GI:12231159"
/db_xref="SPTREMBL:O9EWC1"
/translation="MRTLLVDVDSFTNLPHYLSRANGREBEVIRNDPAVRPGLLD
APDVVLSRPGCTPHRPAADPGICARIAEGGLPVYCVLQCGMALAGARVGRAP
RHGRTSAVRHDSGTGIEGLPOPLEVYRHSLSAVTELPRELETTANSEDCVLMALHRT
LPLMGVOHPESIGTQDGHRLIANRPSLRTGDPHPRMWSRSGEGLQLSMNGDASG
LARTAKADVHAGTVTRADGASSTVESAFITLBNDLAGLRTEVEPLPFAFALGVWG
LGVELKACDGDGAHRSPDPDAVLVFAADRALVDRHTTYLLALVEDDAEABARAL
AAASATLDAVGREREPCPEAPVCTTGVEILRHDDGYLKLIDVCOQRIAGETVEVC
LTNMAEADTDLTPHAAYALRVSAPPAPALIDFGPMALVSSPERFLRIDRHGMS
KEIKGTIRPGATPOEDALVRALATCERDAENLMTVDLVRHDLGRCAVSSVADPV
FOVETVATVHQLVSTVTRAREDSVPAAVRAAPFGSGMTGAPKIRTMQIIDLKGP
RGVYSGAIGVSELTGAVDSLIVIRTVSGRLRYGSGAVIALSDPADEFETAVKA
APLILDTAPFGREAPKDLDGSEDDGTDAGAPDVLVPG"
20829..21836
/gene="canra"
20829..21836
/gene="canra"
/codon_start=1
/transl_table=1
/product="ABC-transporter"
/protein_id="CAC22118.1"
/db_xref="GI:12231160"
/db_xref="SPTREMBL:O9EWC0"
/translation="MNGLLPAPLPAPGVSAAGSDLMIEAGVSKAYGTVALDSV
SLAVRGVTLGLHNGAGKTLVDLITLTPPTSGRARVAGVADKVEIRRRIGL
TGQFASVDAQNSGYNLVIARLDAGRRARVRARELLEPRLTEVADRASSVSGG
LARRLDLAVSLGAREVLFLDEPTTGLPSSIRLIMEIYEGVEGTAVLLTQYLEE
ADRLDLRIAVLSAGRVAAAGTAPELKATVGRRTVTLTETGDEGAARPLAGAPFAP
VDGEARTVVPIDATREIADIIRISDQAGVEASELNFPEPFLDDVYLLAEQAAGRTR
A"
21970..22776
/gene="canrb"
21970..22776
/gene="canrb"
/codon_start=1
/transl_table=1
/product="ABC-transporter"
/protein_id="CAC22119.1"
/db_xref="GI:12231161"
/db_xref="SPTREMBL:O9EWB9"
/translation="MPRAAVPAFGSSLLFTQIRVLVTSLSRAMVTDPIGVFLGIOP
VVIILFVLTOVESKMGMPHPFDGVSLVYLVAVLVDAAGASMGSGVGLVEDOKNGI
VARLSRLPHVPGALLARSIVGLVRSVAOVIVMALMTVIGVSGAABEALASGL
TLFISFSQAFIAGANLRRAREPQNLALVIFPLMTRASSAYVGVADDPRLGAVAS
VNPILVATDATALVDVGDHALPAVVICAVIAVGVIAVRGFRRL"
BASE COUNT 5144 a 14526 c 14888 g 4756 t
ORIGIN
Query Match 48 5%; Score 258.8; DB 1; Length 39314;
Best Local Similarity 67.8%; Pred. No. 4,5e-29;
Matches 362; Conservative 0; Mismatches 172; Indels 0; Gaps 0;
QY 1 GCGTCGCTTGTGTAACGGGTGGGACGGGTGTGTCGGGTGCGGCGGCGGATCTCG 60
DB 35608 GGTGGCGGTCTGCTACCGGTGTGTAACGGGTGTGTCGGGTGCGGCGGCGGATCTCG 35549
QY 61 GCTGTGTGTGTGGGGTGGGGATCTGCTGTTGTGTGACCGGCGGCGGATGCTCCG 120
DB 35548 GTGGCTGAGCGGTGGGGATCTGCTGCTGCTGCTCGCGGCGGCGGAGACGCTCCG 35489

[illegible]

RESULT 7					
AX211733	AX211733	2700 bp	DNA	linear	PAT 06-SEP-2001
LOCUS	Sequence	29 ffrom Patent WO0159126.			
DEFINITION	AX211733				
ACCESSION	AX211733.1	GI:15523945			
VERSION					

SOURCE ORGANISM	Streptomyces noursei Streptomyces noursei
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

REFERENCE 1 (bases 1 to 2700)

Stroem, A.R., Valla, S., Ellingsen, T.E., Sletta, H.V. and Gulliksen, O.M.

JOURNAL Patent: WO 0159126-A 29 16-AUG-2001;

Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHAFAUNA AS (NO) ; SINVENT AS (NO) ; ZOCHKEV, Serey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brutaaet, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valia, Svein
(NO)

FEATURES	Location/Qualifiers
source	1. .2700

```
/organism="Streptomyces noursei"  
/db_xref="taxon:1971"
```

BASE COUNT	324 a	1087 c	876 g
------------	-------	--------	-------

ORIGIN

Query Match 48.2%; Score 257.2; DB 6; Length 2700;

Best Local Similarity 67.6%; Pred. No. 1.3e-28;
Matches 361; Geneomutations 0; Mismatches 177; T-adj 0; Gene 0

Matches 361; conservative 0; mismatches 1/3; indels 0; gaps 0;

by 1 GGGTCGGTGTTCGTGACGGGTGGACGGGTGTCTGGGTGCCCGGTCGCCGCGCATCTG 60

100

Db 1170 GGAACGGTGTGGTGACGGGTGTAACGGGTCTGGGTGGGTGGTGGCCCGGCATCTG 1229

[illegible][illegible]

Db 1230 GTGCGGTCGTCGCGGCGCATTTGTTGTTGACCAAGTCGTTCTGGTGTGGGTGCTCG 1289

QY	121	GGTGGGAGGGCTCTGCGGGCGAGACTGGCCCGCTTTGGGGGGGAGAGTCCGATTTGTTGCG	180
Db	1230	GGTGGCGCGGGGTTGTGTCCCGAGTTGGAATCTGTTGGCGCGCGGGTTGTGGTTCGGCG	1344
QY	181	TGTATATGTGGGAGAGCGCGCGGAGATGATCCGGCTCTGAGAGGTTTCTTCGCGGGTGT	240
Db	1350	TGTATATGTGGGTGATAGCTCGGCTGTTGCCGAGTTTGGTTGCCGGGTGTTCGAGAGCGTAT	1400
QY	241	CCGCTGACGGGTGTCTGTGCATATCGCGCTGTGTGTCTGCATCATTCGACGATTCGCTCTTC	300
Db	1410	CCGTTGTCTGGGTGTGTGCATATCGCGCTGTGTGTGTGATATGACGATGTGTGGGTTCTGTG	1465
QY	301	ACGCGCCGAGCGGCTGGGCGACGGTGTTCGCGCGCACGAGGATATGCGCTCTTTGGCTGAT	360
Db	1470	ACGCGCGAGCGGTTGGCTGCGGATTTGGTGTCTCAAGATGTGATGTGCTGTGAACTTGAT	1525
QY	361	GAGCTGACGCGCGGGGTATGAGCTGTTCGCGCTTGTGTGCTGTCTCTCGGCGCGGGGATC	420
Db	1530	GAGCGGACGCGTGTGTCTGTGATCTGTGAGCGCGTTTGTGTCTTCTGTGTGTTGGCGGATGTG	1585
QY	421	CTGGGGTTCGCGCGGGCAAGGCACTAATCGCGCGGCAATCGCGCTTGTGCATCGCTGTGCG	480
Db	1590	TTTCGGGGGTGCGGGTTCAGGCGCAACTAATGCGCGCGGGTATATGCGTTTGTGAGCGCGTGTATG	1645
QY	481	TACCGCGCGCGCGCGGCTCTTCGCGCGGGTGTTCGCTGTGGGTGTGGGCGCTGTGG	534
Db	1650	GTTATATCGGGTGTGTGTGGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	1703

RESULT 8					
AF420594					
LOCUS					
DEFINITION	AF420594	2706 bp	DNA	linear	BCT 25-OCT-2001
ACCESSION	Streptomyces noursei rearranged polyketide synthase (mgcd48) gene, partial cds.				
	AF420594				

KEYWORDS
SOURCE
.
Streptomyces noursei.

Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS
Brautaset, T., Bruheim, P., Sletta, H., Hagen, L., Ellingsen, T.E.,
Strom, A.R., Valla, S. and Zotchev, S.B.

polyketide synthase gene in *Streptomyces noursei* ATCC 11455

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 2706)
Bratasec,T. and Zotchev,S.B.
Direct Submission
Submitted (18-SEP-2001) Biotechnology, NTNU, S. Saelands v. 6/8,
Trondheim 7491, Norway

FEATURES	Location/Qualifie
source	1. .2706

```
/organism="Streptomyces noursei"  
/strain="ERD48"
```

```
/db_xref="taxon:1971
```

gene
c1. : 52/06
/gene="mycC948"

1/gene: mybcat
 <1. .>2706
 CDS

/gene="ny8Cd48"

/note="NysCd48"

```

/codon_start=1
/codon_end=1

```

```

/clang1_cadre=11
/product="rearranged

```

```

/protein_id="AAL18806

```

/db_xref="GI:16417584

```
/translation="GSADDTG
```

TTVWPRDAERLTIDBCYAH
AAECIHPATDCHNAT

VTADPAGNPVATVTRILARI

LSGPTAVLADALGDAVAVAHATLDDLLAGDTTPATVLPVLCAPLDGPTAOAHALT
RSALTLVQOVLATRLADSLRVFVTHGAVATDAPPTDLAAAVWGLISAQENPGT
FTLLDDEPDSSTLRAALRATLDEPOLLRAGARARLRTAPPTTHTGATPAGT
VLTVGGGGLGGLVARHVLVRSRGRHLLTRSRGVAAGAAGVAELSEAVVAVAA
CVDGSAVAELVAVSVSEYSLSAVVAAGVLDGVTGVTPTPLRAVLRPKDGA
LHEATRTGLDLDAPVAVSVAVGAGQANAYAGNAPDLMAVHRVAGGIPGSLWAG
AMQGVMTAGLTERVRAAESGMPLTVDOVALFDALATGSAVLAIVRDLAL
RTRGDIALPLRLGLVAKAIPRLAAATTPGDTLQALQRLQAEPRDLALVRQAAV
LGHSTGDGVDSPRPAFDLGFDSLAVELRIRIAAGLRLPAVAVDPTADALAVL
LTELGLDABSDPDEPDPAGPTDPIVIGSCAPGPGISPEDMLLGLGAVAV
TDPTPRCHDLDMLXDPDPAHACTYARCTGFLHDADPADPFGSGPREMATDSQ
RLLESSWEATERAGIDPLTLRSRTGVAVMVSIGTRLDAAEF
<1..255
/gene="nysCd48"
/note="Region: dehydratase domain DH4; module 4"
1171..1914
/gene="nysCd48"
/note="result of recombination between KR4- and KR5-
encoding regions within nysC; results in deletion of one
complete module; Region: hybrid ketoreductase domain
KR4/5"
2011..2232
/gene="nysCd48"
/note="Region: acyl carrier protein domain ACP"
2296..>2706
/gene="nysCd48"
/note="Region: ketosynthase domain KS"

BASE COUNT 327 a 1088 c 877 g 414 t

ORIGIN

Query Match 48.2%; Score 257.2; DB 1; Length 2706;
Best Local Similarity 67.6%; Pred. No. 1.3e-28;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1 GGGTCGCTGTTGGTACCGGTGGAACGGGTGCTGGGTGCGCGGTGCGCGCATCTG 60
DB 1171 GGACCGGTGTTGGTACCGGTGGAACGGGTGCTGGGTGCGCGGTGCGCGCATCTG 1230

QY 61 GCTGCTGTGTGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 120
DB 1231 GTGCGGTGCTGTGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1290

QY 121 GGTGCGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 180
DB 1291 GGTGCGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1350

QY 181 TGTGATGTGGAGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 240
DB 1351 TGTGATGTGGAGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1410

QY 241 CCGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 300
DB 1411 CCGCTGCTGCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1470

QY 301 ACGCCGAGCGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 360
DB 1471 ACGCCGAGCGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1530

QY 361 GAGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 420
DB 1531 GAGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1590

QY 421 CTGGGGTGGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 480
DB 1591 TTGGGGTGGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1650

QY 481 TACCGGCGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 534
DB 1651 GTTCATCGGCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 1704

LOCUS AX211705 65140 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 1 from Patent WO0159126.
ACCESSION AX211705
VERSION AX211705.1 GI:15523937
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 65140)
Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Strom, A.R., Valla, S., Billingsen, T.E., Sletta, H.V. and
Guiliken, O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 1 16-JUG-2001;
Norges Teknisk-Naturvitenskapelige Universitet (NO) ; STIFTETSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOOGSKOLE (NO) ;
ALFAARMA AS (NO) ; SINVENT AS (NO) ; ZOTCHEV, Sergey Borisovich
(NO) ; SEKUROVA, Olga Nikolayivna (NO) ; FJAEVRIK, Espen (NO) ;
BRAUTASET, Trygve (NO) ; STROM, Arne Reidar (NO) ; VALLA, Svein
(NO)

FEATURES
source location/Qualifiers
1..65140
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"

BASE COUNT 8270 a 25171 c 22273 g 9426 t

ORIGIN

Query Match 48.2%; Score 257.2; DB 6; Length 65140;
Best Local Similarity 67.6%; Pred. No. 6.9e-29;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1 GGGTCGCTGTTGGTACCGGTGGAACGGGTGCTGGGTGCGCGGTGCGCGCATCTG 60
DB 25919 GGACCGGTGTTGGTACCGGTGGAACGGGTGCTGGGTGCGCGGTGCGCGCATCTG 25978

QY 61 GCTGCTGTGTGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 120
DB 25979 GTGCGGTGCTGTGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26038

QY 121 GGTGCGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 180
DB 26039 GGTGCGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26098

QY 181 TGTGATGTGGAGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 240
DB 26099 TGTGATGTGGAGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26158

QY 241 CCGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 300
DB 26159 CCGCTGCTGCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26218

QY 301 ACGCCGAGCGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 360
DB 26219 ACGCCGAGCGGTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26278

QY 361 GAGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 420
DB 26279 GAGCTGACCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26338

QY 421 CTGGGGTGGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 480
DB 26339 TTGGGGTGGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26398

QY 481 TACCGGCGCGGAGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 534
DB 26399 GTTCATCGGCGGTGCTGCGGAGGTGCGGATCTGCTGTGTGAGCGCGGTGCGCGCATCTG 26452

RESULT 9
AX211705RESULT 10
AF263912

LOCUS AF263912 123580 bp DNA linear BCT 24-MAY-2000
DEFINITION Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
complete sequence.
ACCESSION AF263912
VERSION AF263912.1 GI:8050835
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 123580)
Brautaset, T., Sekurova, O. N., Sletta, H., Ellingsen, T. E., Strøm, A. R.,
Valla, S. and Zotchev, S. B.
Biosynthesis of the polyene antifungal antibiotic nystatin in
Streptomyces noursei ATCC 11455: analysis of the gene cluster and
deduction of the biosynthetic pathway
JOURNAL Chem. Biol. 7 (6), 395-403 (2000)
MEDLINE 20334850
PUBMED 10873841
REFERENCE 2 (bases 1 to 123580)
Brautaset, T., Sekurova, O. N., Sletta, H., Ellingsen, T. E., Strøm, A. R.,
Valla, S. and Zotchev, S. B.
Direct Submission
JOURNAL Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
N-7489, Norway
FEATURES
source location/Qualifiers
1. 123580
/organism="Streptomyces noursei"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:1971"
/complement (46..783)
/gene="nysF"
complement (46..783)
/gene="nysF"
/function="presumably post-translationally modifies ACP
domains on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=11
/product="NysF"
/protein_id="AA071762.1"
/db_xref="GI:8050836"
/translation="MIELLPATVATEAAYDPRPGRLLSREEVIAAVESRORE
FTVRIHARALRLGDPRAIPNRRGAPOMPGIVGSMTCAGYRAAASPALESA
AVSIDAEPNGPLPAGVLAIALPSERPHVALLAARPDVHMDRLLESASVEPKMYP
LQRELDSEAEIVDPTOGAFARLLVFGPLLGRVTVPEGRMHTPALLTTAVHL
PAPTRRDEHRTHTVNSPLPRPTG"
complement (867..2684)
/gene="nysG"
complement (867..2684)
/gene="nysG"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysG"
/protein_id="AA071763.1"
/db_xref="GI:8050837"
/translation="MASPDLEBERTPAPPARLVGLRPHRRSVALAVSGVIGVIL
NAFGLLRVTDLIADGVGVGPAPGIDPAIGRLLVLLVLAIVASLMLQGR
LVAASVWRTHELRDAREKLTPLPHRDPAGELLSRTNDINLOOTLQCTLA
LTSISLTMLVLMVISPLAVMLLSVPYSAIARIKRAOPHYAASANGTL
NAHVEVCTGHALIKGFDRRAAEERFDACNDAYVRAAKAOPASGAMPEVMMVAVL
GVAVVITAMKVIKNTLTLDGVQAFILYARQSPQIVETIASAGLQSGIASAQRF
TLIDAPQAPDPLRGPAPABGRVETVVSFSDPTLIENTSLTVPSGSAIVIG
PTGAGTKTGNILMRFYEPDGGRIIDGDTITMTDDLSRFGVLQDTWFGGTA
ENIAYAGPACRADIEAARACADRFITLPGQYDTVADSGTASAGEKOLLTVAR
AFLARPAVLVLEATSSVDTRTEVLIFORAMNSLRAGTSFVLAHRLSTRDADLVVM
DAGRIYEGTHOOLLCAQGLVARIHAARHTTPAGAAAG"
complement (2662..4416)
/gene="nysH"
complement (2662..4416)
gene
CDS

/gene="nysH"
/function="presumably involved in efflux of nystatin"
/note="putative transporter (ABC family)"
/codon_start=1
/transl_table=11
/product="NysH"
/protein_id="AA071764.1"
/db_xref="GI:8050838"
/translation="MLRLRLRAQRLPYAMATPAVALQVQITGLTLPTGLALIDQ
GVKDGGRITREIVGMGVNVAIVQIAALGAALAAATATATMRGRRLAPLPRITIDS
AREIRFGPSLLITSVNDVQVNLATQGTREIVCAELMCGSLTALRQVPIALL
LVALVVAVCGFLIARNGTIVAMQTLTDLGLLEALITGVVVSFVDDHSLA
RPAQNDNAFLVSRVGRILATMLPVLILMNGFVALIMTSHRIDGRMDEISRA
LSTYSLIMSVYMAFVFLSVPRVACAGAAEVLDTGSSVAPPAAPQVGPAGRI
ELCAGVYRPGAEFVLVDVDTVEPGRIVAVIGSTGSKTILVLIADATGAV
RVCQDVREITPAATLAANGVPRPQVYPSGVASNTLRGPDADEELAEALVQAA
ADPVARMPDGLDAETTQGGGVSGGRDRSLARLLRPETLTLDDEFSALDQATDA
ALRTALVPTAGATVITVAQRISAGRDDRIVLDRGVVAGTHTDILLRTPTYREI
4714..5748
/gene="nysD11"
4714..5748
/gene="nysD11"
/function="presumably involved in mycosamine biosynthesis"
/note="putative dGDP-mannose-4,6-dehydratase"
/codon_start=1
/transl_table=11
/product="NysD11"
/protein_id="AA071765.1"
/db_xref="GI:8050839"
/translation="MSKRALITGITGODSYLAELHLSQGVQVWGLRQANPRKSRV
SLASBLDPIDGLMDQGLSVAVDVQDPDEVYNIGALSFPQMSWQALVTEVNGMG
VLRMLEAIRMGVSLSTRVSPRQIRFYQASSEMFGKAETPORETTLFPRSPYG
AKAGVGYITNRYRESFGMAVSGMLFPHESPRGREGVETRKISLAVRIRQGLDYL
ALGUNDAYRDMKGYADIVYRAMHMLQDAGDDYVYGTQMSVRYAVLAIBHYELNM
EDYVIDPDLVRPAVEVLCADSARAQRLGKRPVDPPTLRMNVDSDLQVSRNQ
YGDVLLAANW"
5930..34363
/gene="nysI"
5930..34363
/gene="nysI"
/function="responsible for condensation steps 9 to 14 in
the nystatin polyketide backbone synthesis"
/note="polyketide synthase"
/codon_start=1
/transl_table=11
/product="NysI"
/protein_id="AA071766.1"
/db_xref="GI:8050840"
/translation="MDNEOKLRDYLKATADLRTRRRHKLKESNAQEPVAILGWTCT
YPGVSPEDLMRWAGEHGVTPPTRGMDLALAAPASGGFLHDAPDFADFG
GISPREAVAMPQORVYTESAMEAERAGIDPTSVKSGRTGVFGAMADQRYVAGD
AEGFOLNTGSSVLSGRISYTFGTGAVYVDTACSSLVAVHALTQLRAGECTAL
AGGVITKSGPFTLEMGRQGLSADGRCSFGDTAGMGAGVGILVLESLDVNRN
GHEILAVRGTAIVNODGASNGILTAPNGSQOQVIOQALVNARAGDIDVVEARHGT
TLGDPEAQLLATYGNRPADRPPLLSGVSNSHSTQALVNGAGVAVIKNWAMRHGT
PRTIAEERTHVDMSOGAVRLDITDTPMTPAGRAASVFGSIVLPLNVTGKPA
EPPEDAATADDDAAGSTPATAPVAPVGPULSGRTPAVDAIRGQAAALRALDTRRP
DILDLASLATTRACFERAVULATDHALIDGUTALADADDPAAPMWTGTTAAET
RLAVLFTQGAQRUDAGREELARFPFAETALDAADATTPHLDRPLRVLEWGTDLA
DRTAAQAGLDAVEVALYRLIESFGVRPDHLAGHSGVEIIVAAHLGAVSLDADTLVA
AGRLMOALPGGMIAIVQASEADVAAPLAGEHDVVAALAVNGSAAVLSGAELVTA
LAEOALAGKTRRLRVHAFHSPLMEMBLDAFRAVVDLTLOPLLPVNSLTGKPA
TVAOLTSADVVDVHVRHVRPADGIDMLARRDITAELELGDGVLSMAOCLDARDA
DAVITPALRAGPSEHTITTLALGLHNGATLDWTGCGPAGARTDLPYAFQRRY
MFKALQSTADLRSGLEAAHPPLISAASVLSADAGTLLRLSRQWTPMLADTRVG
TTLPGTFLBLAVARAGDEVGCDRELETLAPLILPQGGVQVQLMGNPDVSGRT
VNVHARPTGQDPTATATGVLTADLSRLPAPSSBEGGFLAGDPLMADPAQMP
AGAEILADHGAAPRAGDGGPGVPGVQALRAMGGVAVVLEVELEPAGSDAAEFLG
HPLILDAIDLAAPRAGDGGPGVPGVQALRAMGGVAVVLEVELEPAGSDAAEFLG
ADGTGAPVLSVLSVATQOQDITAAVARDALFRIDMTVPVQPTATDPPGVALIGA
DPFGLLTAGRADAPADVADLALAAADGPPVTVVLSAGDGDADPARSAHRAAE
ALAAVQTWLIDHHERFAARLVFTRGATVGRDVAAAVWAGVLRSSQSENPGCFALVDL

DPGAVGAALVAALVSGEPQLAVRGDVLKVARLVRRPLTEVGADGTGDDGDSG
VSPSGGAVLVTCGTGGLGAVLRLVAEYVADLLVSGSGRAVAGABLVRLVGLG
GAVRVACVDRRAVELVGVAVSAVVAHAGVADDDGERSAVLAPYD
AMHLEATRGDLDAFVSVSSLAGVSGGQANVAANAFTDTRRABRLPGLS
LAMPWEGSGMTCTLDVDAERLARGVPLSVAQGLAFDAVAGTCAVCPYLD
LPLRARGVPLRLSLIRVARRAVAGSALGNLAQRRLDEGRDBMDLVLRG
QVALVLGATGSDVADGARFRLGFDLSLAVELNRLNTVGLTALTYLFDVTRH
LATYVLDLDELTAEVATAVPAVADPPIVVGACRPGVSGSPEDIMVLTGTG
GAVGPTNRGMVESLYHPDPHSGSYRSGEPGSEAGEPFGSGPRALTDG
SOORLLESSEMEI ERAGIDPVLRSRGTGCVGAWYSDVSNALAPREPGSGSS
PILASGRVATLLEGPATVTPDCCSSLYAMTMOMALRSGCGALAGCTYMTSP
AVFDRARQKLSPEDECKAFADAGVSGVGVLEVRSDVARNGHEILTEQA
SAVNOGASNLTAHPNGSQVIRQALASGGLADVDENHAGTGLTDIEQA
LATYGRDREPERPLLGYSKNGHTQAAVAGYIKWLAHRHGVPRTLTADP
SHVDEGAVELLSBOAMPETGRVAGVSGFSGTNTVVEQAPALAAAGA
ARRTPAPVTLISGRSALRGQARLGLHQRAPALVDVLSLATTRESQOAA
VVAQDRQLASLGLAALADPRPAVVEGAAGRRVAVFTGCGSORAAMGREHEVO
PEFAAPFADCAVFDPLDRPLEVVAEEDGSEALBDETGTQALPRAVEALRVO
VESMGVPRDFVAGHSIGELIAAHAVGLTLEDCRLVAANAATMOMLPFGAMIAIOA
TEDEIAHLDDTVLAAVNGPQSVISGDEEAETIATPFAERGRKRLRSHAFS
PRMDGLADARIVAEGLTYRAPRIPLVSOLDGRADAEVCTGAEVVRHREVRAD
CVTLRDAGATFLELSDGLTAMAEITLGGPIRQPIAEVPMIABGAELAAATLAR
LOVRGVDMMAIAGTARTRDLPYAPGHATYRQLPRAALAAAPADQMLAA
VERGDARLADLIGLGEQDLTLPDLSLPLATSRKNGOEHLDTLRVEMWRLSKP
TARVLDGTWMLVSDATAADQPALDGLDALGSHARVRLLDSDCADRAVALERL
ARTADVAATPOTVGLPLDERDADCPRLTRGALVVALQALADTCAOGLMTATRG
AVETNPADPTVHVOAAMGLGVALHEHRLGKGLVQVDEBAGAGLADILVVK
DADGDOVALRATVSGRVLVHTVETALPTAEPFATVGLITGGTGIGAVANWL
ARAGOHVLTSTRGDPARGAELRABEYSGSVVACVADRALAAVLTALDEE
LPLTGVHTHVGHIGLDTLSTAEFRAGLTAALGAHLDLALARELDLFLVPSI
AGWGSQNSAGANAVYLDALALHRRARLAITSVAMWGMAADVAADAVETLR
OGLGLDPAFAMTELRAVQVTVTVADVDQVAPLPTSPRALLAGLEVERAL
AABEREDOTAGSEVTVRVALRABEOLLELTDLVBESATVGLSGVAAVEGAF
ROYGPISTLVELRKLGAATGLSLSTPEVDPYPTLEAOVRAILICAVLEAVP
ATCGADDEPAILIAGMACRPGVSSPEQLMDLYASTDAISEPVNRGMOTGLPPD
PDRPSTLSTGGFLEHDEDFPTFGLSRELA.VMDPOQLLETTEMSFEAGLRD
ETLRSTLTFGGFSSYQEVGLGADGTEGMWGVSSPGLSVFGLFEGEAVTAD
TACSSILVALHLACSLRNESMLAVAGATITTPNPFAERORALADNCGCKFS
DDADGTLAEGVVLEVERLNRONGHPVLAIRGSAIINDASAGLNLTPNSGOR
VIRORALANARIPGIDIDALEHGTGPIGPIEAKOALFAYGSDRDEPESALLSVKS
NIGHTOSAGIASIVIKMAKRSELPPTIHAADPSHVMSAGTRILTOANREBT
GRPRRAVSSFGISGTNAVHLEOAPADTPAERPAERPAVPAAGVENVVARSA
ALRGOERLAAHEVGTALPAAGPLDIGLSVARSARPEPVAVVPFAETDIALR
AVATDPSPVARAGVADVEGRVFPVPGSGQWVGMSQLDSDSAFABEACAL
AEFTDMSLVNLRGVGAPSLERVDVQSPAVMYSILAMSRGLPRAVGVHSGV
EIAAVVSGALSIDGARVVALRSOALGRALRGMSVSLSDVLEPLVPEBRV
SVAAVNGPRSVVAGEPEBALDHALHARLTADIRAKIANDVASHSHOVEDHELELV
LAELAPTSSEVPFSTVTDMDLTARMDAGYFRNLRGRVFAVADAVALAAEYRAFV
EVSSHPLSMANVOEALIDEAGVPAAVAGTLERDQGTDRPLLSAEVFRGVDMAGL
FEGTGSARILDPTVAFQHEHLMVPPAPPEAVAADPDAAFMTAVEGDVSLTALAG
TDEDSVAVALPALTSWRBARRDRSTVDAMKVRVAKLGGTLPLHPSLGTWMLVTADG
IDPDTVAAGLETYGAEVRRLVDEBCDRAVLAERLAGADYTGYSVLAARRTAV
PGLSVLGTALVIALTOALGDABEIDAEVAVLGTGAISTGADLRLPVQAOVGTICHT
AALEHPQRMWGTLDLPAALDARAQAORLAVALSGLSDQDLAPFGVPRTRIVRAEA

Query Match 48.2%; Score 257.2; DB 1; Length 123580;
Best Local Similarity 67.6%; Pred. No. 6e-29;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGAACGGGTGGAGCGGGTGTGCTGGGGTCCGCGCGCATCTG 60
DB 85772 GGAACGGGTGTGGTGAACGGGTGGTGAACGGGTGTGCTGGGGTCCGCGCATCTG 85831
QY 61 GCTGCTGTGTGTGGGGTGGGGAATCTGCTGTGTGAGCCGGCGTGTCCGATGCTCCG 120
DB 85832 GTGCGGTGTGTGTGGGGTGGGGAATCTGCTGTGTGAGCCGGCGTGTCCGATGCTCCG 85891
QY 121 GGTGCGGAGGATCTGCGCGCGAGACTGCGCGCGTTGGGGCGGAGAGTGCAGATTGTTGCG 180
DB 85892 GGTGCGGCGCGGATGCTGCGCGAGAGTGTGAGTGTGGCGCGCGCGGTTGTGTTGCGCGCG 85951
QY 181 TGTGATGTGGGGAGCGCGCGAGAGTGTGCTCCGCTGTGAGAGGTGTTCTCTCCGGGCTGT 240

DB 85952 TGTGATGTGGGGTGTGATGCTGCGCTGTGCGAGACTGTGTCGCCGTGTGTCGAGATCTGAT 86011
QY 241 CCGCTGACCGGGTGTGCTGTGATCGCGCTGTGCTGTGTCGACGATGCGACATCCCTCTC 300
DB 86012 CCGTGTCTGTGCGGTGTGATCGCGCTGTGCTGTGTCGACGATGCGATGCGGTGTGCTG 86071
QY 301 ACGCCGAGCGCGCTGCGCAACGCTGTGCGCGCCAGAGTGAATGCCCTTTTGTCTGAT 360
DB 86072 ACGCCGAGCGCGTGTGCTGCGCTGTGCTGCTGCGAGAGTGAATGCTGAGAACCTGAT 86131
QY 361 GAGCTGACCGCGGATGATGAGTGTGCGGCTGTGCTGTGCTGCTGCTGCTGCGCGCGGATC 420
DB 86132 GAGGCGACCGCTGTGCTGTGATGACGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 86191
QY 421 CTGGGGTCTGCGCGGAGGCAACTACGCGCGGCAATGCGCTGTGAGCGCTGCGG 480
DB 86192 TTGCGGGGTGTGCGGCTGACGCACTACGCGCGGCAATGCGCTGTGAGCGCTGCTG 86251
QY 481 TACCGGCGCGCGCGCGCGCTGCGCGGGGTGTGCTGCGCTGCGCGGCTGTG 534
DB 86252 GTTCATCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 86305
RESULT 11
AX211739 125401 bp DNA linear PART 06-SBP-2001
LOCUS AX211739 Sequence 35 from Patent WO0159126.
DEFINITION AX211739
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei.
REFERENCE 1. Ibañez I to 125401)
Zotchev, S.B., Sekurova, O.N., Fjaervik, E., Brautaset, T.,
Strom, A.R., Valla, S., Sillingen, T.E., Sletten, H.V. and
Gulliksen, O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STITTELSEN
IND OG TEKNIK FORSKNING VED NORGES TEKNISKE HOESKOLE (NO) ;
ALPHARM AS (NO) ; Sivent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES
source location/Qualifiers
1..125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 48.2%; Score 257.2; DB 6; Length 125401;
Best Local Similarity 67.6%; Pred. No. 6e-29;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGAACGGGTGGAGCGGGTGTGCTGGGGTCCGCGCGCATCTG 60
DB 86180 GGAACGGGTGTGGTGAACGGGTGGTGAACGGGTGTGCTGGGGTCCGCGCATCTG 86239
QY 61 GCTGCTGTGTGTGGGGTGGGGAATCTGCTGTGTGAGCCGGCGTGTCCGATGCTCCG 120
DB 86240 GTGCGGTGTGTGTGGGGTGGGGAATCTGCTGTGTGAGCCGGCGTGTCCGATGCTCCG 86299
QY 121 GGTGCGGAGGATCTGCGCGCGAGACTGCGCGCGTTGGGGCGGAGAGTGCAGATTGTTGCG 180
DB 86300 GGTGCGGCGCGGATGCTGCGCGAGAGTGTGAGTGTGGCGCGCGCGGTTGTGTTGCGCGCG 86359
QY 181 TGTGATGTGGGGAGCGCGCGAGAGTGTGCTCCGCTGTGAGAGGTGTTCTCTCCGGGCTGT 240

gene
CDS

```

/db_xref="taxon:1906"
1..43280
/gene="tylG"
816..1434
/gene="tylG"
/note="ORF1; putative"
/codon_start=1
/transl_table=11
/product="tylactone synthase starter module and modules 1
& 2"
/protein_id="AAB6504.1"
/db_xref="GI:2317860"
/translation="MSSALRAVQSNCGYGLTNTAQTQDQEDVDGPDSTHGGE
IIVGMSCLPGAAVEEFMELLRSGRGMTRDDGTWRALIEDHAGPAGFGMMAR
QAAATDPORHMLTEIGMEALBDAGIYGLDTGDPVPAVLSDDAVLTTRSSVAG
GYATGTHRALANRSLHGLRGPSLYVDSQASALVAVQACSLRGRTSLVAG
GYNLLTEESTTWMERWGLSPDGRCHTDARANGVYRDEGAVVLRFLDAALADG
RVCVTKGAANNDDGASLITPDREAQAVLROAYRRAGSTGAVRYVELHGTGRA
GDPVEAALGAVLGAAGDSGRSTPLAVGSKTVNGHLEGAAGIVGLIKATLCVRKEL
VPSLNFSTNPDIPLDRLRQTEROENNEEDPRVAVSGFSKMGQTNHVLVIAEA
PAAAGSSGAGGAGSAGISAVSGVPPVVSGRSRVVVREAAGRLEAVEAGVGLA
DYAVTMADSRRCRYAVYLARGAELAGRLRAGDPPAGVYTGAVLDGVYVGAAP
GGAAGAAGAGAGAGGAGVLPFGQGTQWGMGGLIGSSVFPAISRBECKALSY
HVGWDLLEVVSAGGLERDVVQPVYMAVWSLARWQMGVDVAAVHSGGEIAA
TVAGALSLDDAAVVALRAGLIGRYLAGGMAAVPLPAGVEAAGLAKMPGVEAVAN
GASIVSGDRRAVAVYAVCOAGEVQARLIPDYASHRHYEDLKEGLERYLSGIRP
RSRPVVCSTVAGEOGEPEVPAGVFWRLNRNVERSAVVGGLIEEGHRRPIEVSHP
VIVHATEQTAADRSVHANTGLRRDDSPHRLTSTAAMAHGATLTMDPALPREHL
TTLPYTFNNHHVMDITDGGDDDTQEKESQPLRELRGLSPSKQGLJLDLVCRA
HTAVVGLDTAAEVMDPLSPKQGIQSMTEVLRNLLETGLALPTLVYRPIERA
LAQFLTELLDGSPPSGVLAPOKSEAOEPLAVGMGRFPGVSGPALMRLVVEG
VAVSPFPDGRMDVEGLDPEBGVAGKSYREGEFLDAEEDAEFFGDISREAVAM
DQOORLLTSMETIERAGIDPHSLHSGRTVAGVMPQEGRLAEGAGSDGYLT
GTSGVSGRVAYTLLEGPRATVDTACSSLYALHLTAQALRGCDMAALGCTVM
AGPKNVERSRQGLADGRCAFDAGAGITMAEGAVLVERISDARLGHPLVAV
VCGSALNQDASNGLTPAGSPQERIRQALGNARLTVDVAVEHAGGTLRGDIE
AQAALLCTGRDRDGRPWIGLSKSNIGQAAAGVAGYIKWLMRYOMLPRTLHVD
BPSRHYDMSAGVWLTEARWMPGVDRPRRAVSAVAGSTGANHLILEADTAEASA
TTPVRESESSAAVLDARSAGVPPVVSGRSRVVVREAAGRLEAVEAGVGLADAVT
MAGRSRGYAVYLARGAELAGRLRAGDPPAGVYTGAVLDGVYVGAAP
PQOGTQWGMGGLIGSSVFPAISRBECKALSYHVEMLLEVSGAGLLEAVDVP
VTMAVWSLARWQMGVDVAAVHSGGEIAAIVAGALSLDDAAVVALRAGLIGR
YLAGRGAAMAAPLPAGEVEAAGLAKMPGVEAVANPASIVSGDRRAVAVCOAG
GVQARLIPDYASHRHYEDLKEGLERYLSGIRPERSRPVVCSTVAGEOGEPEAG
YWRPNLRNVERSAVVGGLIEEGHRRPIEVSHPVIVHATEQTAADRSVHANTGLR
RDDSPHRLTSTAAMAHGATLTMDPALPREHLTTPYTFNNHHVMDITPTPAT
TTOSPTDARIRYTKALTESSEFVRPHSIGRCLVAPPTDGLDGLTTLVLSERKAS
VARLEVPIGARRAEVALLKPSMESAGEENTVVSILGVPSTDAVRTSIALIQVSD
IGVPAARVVALTRRAVAVCGETPDAGAQMGFVVALLEPDIWGGLIDIDPETHEL
TRPETSOPOTPERLPOTNRRLALEAAVLAGRGEQVAVRASGIGRAYSRAA
AQAASWQSGTVLITGGMGALGRRLARLARGAERLVTSRRGPAPGAALAEELR
GHGCRVYHAAQVABERDALALVTAAPPNAVHTIGLIDDAVITLSPSEFTVYGA
VCGAELLHOUTADIKGLDAFVPSSTGTWAGNAGVAAANAALDALERRRAGLP
ATSVAVMGGGAGAGAGESLSRRGLAMPDAVALLAMGNDVAVCVVVDV
ATSVAVMGGGAGAGAGESLSRRGLAMPDAVALLAMGNDVAVCVVVDV
VBLVRETAFLVLRHPNTDAIGARPEKSGFDSLTSELRNLNAGTGLKATVIFD
HSPPTARLILLDLRLTGAGAPAPADEPRLPAVADDDVAVVVGMACRPGAGPBEA
LMLKTYBERDVGAAPTDRGMDLSYDEPREPAGCTTYRREGFLHDAEFPABEFGI
SREAVAMPQORLLETSTWEIERAGIDPHSLHSGRTVYVGLTHQETASVLEHEP
EXEGYLLTKSASVSGRISYTLGLEPGLSIDTQSSSLVLAHNAQAALRGECMA
LAGVYVMAAPGLFVEFSRQGLADGRCAFDAGAGITMAEGAVLVERISDARLGH
LGHPLVAVCOAGGVTVGAVVDEBETSGGCGGVLPFGQGTQWGMGGLIGSSVFPA
SMRECKARALSVHENDLLEVSGAGLERVDVQPVTAVMVSLARYOMGVDVAAY
VGHSGEIAAATVAGALSLDDAAVVALRAGLIGRYLAGGMAAVPLPAGVEAAGL
KMPGVQVAAVNPAGSTVSGDRRAVAVYAVCOAGEVQARLIPDYASHRHYEDLKE
ELERVLSGIRPSPRPVVCSTVAGEOGEPEVPAGVFWRLNRNVERSAVVGGLIEG
HRRFLVESHAPLVVHATEQTAADRSVHANTGLRRDDSPHRLTSTAAMAHGATLT
TMDPALPREHLTTPYTFNNHHVMDITPTDAGVGAAPRFGMTWEDHPLRGLPLA

```

```

misc_feature
942..3929
/function="encodes starter module"
/gene="tylG"
942..2156
/function="encodes beta-ketoacyl ACP synthase KS-Q of
starter module"
/gene="tylG"
2571..3557
/function="encodes acyltransferase domain of starter
module"
/gene="tylG"
3675..3929
/function="encodes acyl carrier protein domain of starter
module"
/gene="tylG"
3993..8471
/function="encodes acyl carrier protein domain of starter
module"
/gene="tylG"
3993..5264
/gene="tylG"
5631..6617
/function="encodes beta-ketoacyl ACP synthase of module 1"
/gene="tylG"
7410..7949
/function="encodes acyltransferase domain of module 1"
/gene="tylG"
8220..8471
/gene="tylG"
/function="encodes acyl carrier protein domain of module
1"
8541..13970
/function="encodes acyl carrier protein domain of module
2"
8541..9812
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 2"
10260..11246
/gene="tylG"
/function="encodes acyltransferase domain of module 2"
11319..11876
/gene="tylG"
/function="encodes dehydratase domain of module 2"
12861..13415
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 2"
13719..13970
/function="encodes acyl carrier protein domain of module
2"
14351..19945
/gene="tylG"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"

```

```

DSGERVFAGRLAGSEHMDLTDHAVSGVTLDPGTAFAVERFALHAGAATGCGRLBELSVEA
PLVLPAGAGVROMVMSAADSGRRVLIHSAPEAAVHSAAGSGAGVWTGEGTL
VDPPEPTPPDAMWAMPAGRVERGAPVLYERFGLAGYEAPAGVAVWMPQPAL
AEVILPDASTAGRGFVHRPALLDALQPMVINGGLIEPDEDAVILPFWMOGVSATAG
AGALKRLTKADGAVSLQADTSAAVLSJALWRLPLARKLVLDLGTIDGERSLY
RVEMQRLPLPAGPPRMAVLPAGDARLGLTPOGQPGPPALPEVVALRKLALAG
APRPAVAVLPVLSGAGATPESVRQTEECITLQDMLEAEVDPPLVLTGGAVAAY
PGEIIGDALGAVMGVLRASERHGRPALVDTGHPDDRTALPLARVLDGAGOLS
LRAGARTPVLIRACTPEORGPAPDPAGTVLVGATGLGRILARHLAAEHGVHLL
ILSRGRARAEGBDELAELAGLEAPRCPACDAADRELAIVLAEPADRLTGVTHA
AGVLDGTDLDLTERIGVMPKADALNHELTRISPLSVFVPSGAAGILGRPGQ
ANYAANFTFLALQARRAHGIPAVSLMWGMGATGWTGHLSTDLRMRSGLIAPM
THDQALFEDRLALASADPLIVPRLDLALVRRASHGPDVAGPILGLIPAAVAV
ROAAVPGVGAAPAGCGGTARLAGEAALRELVRLVRAVEGVLGYSQPDVAP
GPPFDLQEDISTLAVELNRRLGAATGLRPTALVDFDRPTSOAVAEVLAELAGPRDG
DRAAARFGBELAAVAGLAEDDLRRVLRRLTELAAALTPOGRNSAPAPASDL
DERLSDANDDLFAFIEBOL"

```

```

/db_xref="GI:2317861"
/protein_id="AB06505.1"
/translation="MTAEIVATRKRLGALREARPIAVAVMSCRYGVTTPEDLMR
LADDERDAVSGLRDGMIDLALYDPDGGSTGYARGGTSLACAFDEFEISRE
ALAMDPQULLLETSWELERAGVTADPRAGSTGYAVCMYDNDYARVLXGAGAPP
EDIEGLVNGSASISAGSVSTFGRLGRPAVNTKCSSGLVSLHVALRNLNGECDM
ALAGACTVSTPTVLVDFSRKGLAPDRGKARCAADGSMVAGGMLLORLSDAR
REGRPVLAIVRGSAVQDASNGLTAPNGAQOVIEDLRDAGVPDDVDVEAATG
GTELDPIREGALLATYGTARLTAERPLMGLSLGSHQDPAAGVAVGIMVLAHMG
LBPRLTHVBPRTTAVDMKGVRLTEPVMPEABEPBAGSSVSGASTGAVLVLE
SVAPGPPAAGREDTGAMTYSGRPALRAOALYDALITGTCTGCGAQCQAGP
GTAIVAGALAHATARKHRAVVLGSRNLELALRELAEBHHPVAVGTGAPATER
TAFLESGQSGRQSGRLYRRHPFARALDEVCALBPHLPDLDFEABESPEA
LVDRTETFPOLPALQTLFLAEHNGRAELCHSVEIJAARAGVLTLPDAR
LVARAGLQALPAGAMALRAEAEIAPLERRAGELALAVNPSVAVSGDRA
VIELBOMAREGEARLAVSHAHSRPMQDITOPDRAVLTFTPTPLVSTLTG
TPTRETTCTADHMRQAREPFRFLAMRTLRADGIDTVELEPDDVLSAMRDCA
RPGDITGAGDSTPDLPLPLRLRSVETGHEHPPGERALATVAVHVPRLAP
APDAALAAVAELPTAFTQTHYMLDAPAPALPAGLDAGHPLISALDLEPGRG
VMTGALSAATLPMADHSYGRVLTGTLALDLALHAPVGLTPEADLVLEPDE
RLRVLAEPDASGVRLEISVAGEDGMTSHAVLDGTTTAGEBPAGAPAPAMPD
ABPLDLAEYERFADAGIGYPAFRGLRSRMRDGLIADVRLPGLAEADRYGHP
ALIDALHTRAAALGSAHGMPTTNGVTLMAGAAIRVRLTPAPDVAVYAVDPA
GRVPTAAS/TLRPVTTGQTLAAEARAPLIRVMTGLDPDTGARDHTWAVAGPDL
LPEGTPHHPDLASALADGTAPPRVADLRGVTATPRELASQALAVQMDAAEA
EGRLVLTERRAVDIGDVTDPAAATVAGLVRAQSHPGFALDTPDSRPSROLPR
VAGTAEOLARLDGTLTALPLTRATLPAGARLPALDGTVLITGTSGLAEARHLYTR
HGRARLLTTRSGRPOA PGAEI VAEI ALGAHADVACVADPAARLALIDRHPACH
LTVLHTAGVLDGVTATQTPORLAAVLRPKDAVYRNLHETGHALSLFTYSAG
VLSAGQSGYAAANAYLDSFAPVWRKSRGPAVLSGMGPDSAGMAGLGTDFARLR
SGIAPLSRAEGLALDPAALMAGDDTAPALPLIPRDAVTLRDAVTPAVLRDLACTA

```

```

RESULT 14
MTV018/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
REFERENCE
TITLE
JOURNAL
COMMENT

```

```

MTV018 53450 bp DNA linear BCT 03-AUG-2001
Mycobacterium tuberculosis H37Rv complete genome; segment 90/162.
AL021899 AL122456
AL021899.1 GI:3242282
Mycobacterium tuberculosis H37Rv.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 53450)
Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S. V., Eigmeier, K., Gao, S., Barry III, C. E.,
Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Comnor, R., Davies, R., Devlin, K., Felwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornby, T., Jagers, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A.,
Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skellon, S.,
Squires, S., Squires, R., Suleiron, J. E., Taylor, K., Whitehead, S., and
Barrell, B. G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (6685), 537-544 (1998)
98295987
9634230
2 (bases 1 to 53450)
Parkhill, J.
Direct Submision
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 20, 1998 this sequence version replaced gi:2896750.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original comid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in Tbpasee (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atag, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1..53450
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
51045..>53450
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1..51145
/clone="Y63A"
/feature
/note="fragment designated v018. Does not represent a
physical clone"
complement(114..1370)
/gene="Rv2015C"
complement(114..1370)
/gene="Rv2015C"
/note="Rv2015C, (MTV018.02c), len: 418. Nearly identical
to Mycobacterium tuberculosis Rv1765C(MTY28.31c, 378 aa),
anorf starting next to 1589, and ending in 156110.
Different N-terminus chosen and C-terminus differs as that

```

```

FEATURES
source
source
source
misc_feature
gene
CDS

```

of MTU18639.jlc has been truncated by IS6110. Does NOT show similarities with transposons. BLAST hits with non-IS part of MTU18639. FASTA scores: 295890/MTCY18_31 (378 aa) opt: 2417 z-score: 2553.1 E(): 0; 97.8% identity in 364 aa overlap. Tblastx score is 0.939"

PLAPARIDLTALRAHVAVPMFSDLASAPTRROVDVVAANKSKALAHRLHJLPE
 ABOHAVILGLVRLHIAVTLGNITPEAIDPKAFODLGFPSLITVENRNLKSTGSL
 SPLIFDPTNPLASAIYRTTELAGELOEIPKPAVTPADTPIATIGMCRKYGNS
 PDMMLMLOGRDVLSBPADRCMDLACLYNPDMDLACCTGTPGVGVGDPDPAF
 FGVPSEBALAMDPOQKMLBELSWELEBPAGIDPTGIRGSACTGTPFAVMTQCYMPAB
 PVGFPLTQGLSSVAGRAVYVLJGLEBPAGIDPTGIRGSACTGTPFAVMTQCYMPAB
 LAGHVNATPTVFEFESRHRGLAPDGRCPVAGRADGVMSGGMLVORLSDRR
 LAGHVLAVVGVSAVODGASNGLTAPNGSPQORVVAALANALSLAEVDVVRGHGTG
 TTLDPIEAOALLATATGODRGEPEEPMLSVKSNMGHQAAGVGVIKMVLAMIEE
 LIPATLHVDPSPHVDMSAGAVELTAPRWPAGATRRAGVSSPGISGNMATHIEA
 VPVPRRERAGMAGPVVPMVYSAKSESALRQARLAAVYRGDGLVDVADYMSLARS
 VFEHRAVVGGRDRLLAGLDELADQLDGLSVGRATLAKGTVPFQGSQWMLGNG
 MGLHAGYVPVFAEFNTVYGEIDRLHLRPLAEVWGHDEMLNSTEPAOPALFAVEVAL
 FRLGSGWVRPDMVMGHSIGELSAHVAAGVLSLENAVVAARGLMOALPAGAMVA
 VOAEEVRLPILSAEVDIAVNPASIVISGAONAAVAADOLRADGRVHOLAVSHA
 FHSPLMDPMIDERRAIAAGTIGRPTTIGVSNVTGOLAGDDRGSAVYMRHITFOARF
 ADSVRFQAAGSRFLIEVPSGGLVASIESLPDVAVTTMSALRKDRPEPATITNAVA
 OGFTVGMIDLWRAVVEBAQFVELPTTAFQRRRPLSGDVAADAGLGLASEHALG
 AVIDLPAAGGVVLTGRLSPSVQMLADHSVAGVTIPPGAGFVELAIRAGDEVCGYVD
 BLTLAPLVLPASGVAVQVYVNGPDESGRGVSVSRGVTGWTILAEGLRAGSA
 EPTADLAMPMPAGAVPEVADGYQOLAEGRYGPAPRGLTAMMRGDEVFAVALPA
 DGVSVTGGVHVLLDAILAHAVLSAESAERGOSVLPFMSGVSILHAAGSAVRA
 RIANPVPASVISTELADGLPULSVASMLARPTTDOQLAAVSSSGPDRLPFTYNSPO
 PSAVREPLVPCAGTTEDSAVVFESVPLAGDVAGVYATSSVLDVLSQWLTROGAG
 VLVMTREGVALPGEEDVTLDAAGAAVWGLVRSAGTEHGRIVLVDSDAPLDSSALAV
 TTGEPOVLMRGEVYARVHSGRAVGLLPPSDRPMRLAMSTAGFETSLINKSFAYG
 DAPLPGOVRVAISAIANFRDVMIALGLYPPDPAVWGVACGVETLSTINKSFAYG
 DRYMGLPPEGTGVASTDORLIVKVPAGMSHTAATTSVVFATRAYALVILAAARSGO
 RULIHACTGSGVMAAVOLARHLGLVFPATRSKKNPTLRAMGPDHIDSRSLBED
 KRPAATGKRFVDVLDLAGEFYDASLRLVAPGVFLPEMKTDIRDGVIAQOYRGR
 YRAFDLFEAGPDRIAQILAEATLFGDGVRLPVTTFVDRCPALRLYSQARHTGK
 VMLMPGSWAAGTVLITGGTMAGSAVARHVAVRHGVRLVLSRGRPDAPGAABELVA
 ELAAAGAOVUVACDAADRALAKVADIPOVHLSGVITHAGLADDAVMSLTPDRV
 DVLIRSKVDAAMHIELTDLVSAFVMSMAGIVSSSGOANYAANSEFLDALAHR
 RAHGLPAISLGMKLMDOASMTGSLATVDPKPARADGIYAMSADALQFPTRMATYDE
 PFMUPHIDIPALAKVFDGTLPEPMFVDLIMAPTRROVDLSLAASKSALQORLGL
 PEDBOHVLIDLVRSHIATVLSGASPEALIDPDRAPQELGDFLTAVEMRNLKSAJGL
 ALSPTLIPYPNASALAGYMRRELSSPDTSDAVVAGAEILQRIVASIIPVKRLROAG
 VLDLILALNTEFTSGODPALAPTAEOIADMDLJDVLVNAAFRNDDE"

BASE COUNT 3023 a 6045 c 4734 g 2168 t
 ORIGIN

Query Match 46.7%; Score 249.2; DB 1; Length 15970;
 Best Local Similarity 66.7%; Pred. No. 1.3e-27;
 Matches 356; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1 GGGTCGGTGTGGTGAACGGGTGGACGGGTGCTGGGTCGGCGGTGGCGGCACTCTG 60
 DB 10750 GGCACGGGTGTGATCAACGGGTGGACCGGGATGGCGGGTTCGGCGGTCAACGTG 10691
 QY 61 GCTGTGTGTGGGTGGCGGATCTGCTGTGTGAGCCGGCGTCCCGATGCTCCG 120
 DB 10690 GTGGCTGTGATGGGTGGCGGATCTGCTGTGTGAGCCGGCGTCCCGATGCTCCG 10631
 QY 121 GGTGCGAGGGTCTGCGGGCGAGCTGCGCGCTTGGGGCGGAGGGTGCAGATTGTGCG 180
 DB 10630 GGGGCTGCGGAGCTGCTGCGCGAGTTCGCGCGCGCGGTGCGCAGGTATGCGCC 10571
 QY 181 TGTGATGTGGGGAGAGCGGGAGGTGTCTGGCTGTGAGAGGTGTCTCTCGCGGGGT 240
 DB 10570 TGTGAGCGCGCGGATGAGCGGGGTTCGCAAGGTGATGCGCGATATTCGCGATTCA 10511
 QY 241 CCAGTACGGGATCTGCTGATGCGGGTGTGTGCTGAGACGATGCGACGATCGCCTCTCTC 300
 DB 10510 CCATTGTGGGCTGTGATCCACCGCGCGGCACTGACGACGCGGTGTGATGTCACTG 10451
 QY 301 ACAGCCGAGAGGCTGGGCAAGGTGTGCGCGCCCAAGTGTATGCGCTCTTTTGTGAT 360
 DB 10450 ACAGCCGAGATCGGGTGTGATGTGTGTGCGGTTCAGGTGAGCGCGGTGCGACCTG 10391
 QY 361 GAGCTGACGGGGGTATGAGCTGTGTGCGGCTGTGTCTCTCGGCGCGGGGATC 420
 DB 10390 GAGTGTACTGCGACCTGATGTGTGTGCGGCTGTGTCTCTGATGCGCGGGCTG 10331

QY 421 CTGGGGTGGCGGGGCAAGCACTACCGCGGCAATGCCGCTGTGACCGGCTGGCG 480
 DB 10330 GTCCGATCTCTGGGCGCAGCCCACTATGCGGCGGCCAATTCGTTTTTGGATGCGCTGGCC 10271
 QY 481 TACCGGCGGCGGCGCGGCTCTGCGCGGGGTGTGCTGCGCTGGCGGCTGTGG 534
 DB 10270 GCCCAGCGGCGGCGCCATGCGCTGCGCGGCAATCTCCCTGGGCTGTGG 10217

Search completed: June 17, 2003, 17:55:51
 Job time : 1038.31 secs

DR WPI; 2001-267102/28.
XX New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT
XX
PS Claim 7; Page 92-102; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (ii); (iii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhannose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (ii) at significantly increased levels or produce new derivatives of
CC (ii). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine and trimethylrhannose biosynthesis.
XX
SQ Sequence 29736 BP; 4401 A; 10346 C; 10080 G; 4909 T; 0 other;
Query Match 44.9%; Score 239.6; DB 22; Length 29736;
Best Local Similarity 65.5%; Pred. No. 8.1e-36;
Matches 350; Conservative 0; Mismatches 184; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGGTGACGGGTGGGACGGGTGTGCTGGGTGGCGCGGTGGCGGCATCTG 60
DB 3604 GGCACCGTGTGTCACCGCGGAAACCGCGCGGCTCGGGGCCCTGCTCGACGCCATCTG 3663
QY 61 GCTGGTGTGTGGGTGGCGGATCTGCTGTTGGTGAGCGCGGTGGTCCGGATGCTCG 120
DB 3664 GTCAGGCCCAAGGTGTGGGACCTGTTGCTCGCAAGTGGCGGTGTGGCGGCGCC 3723
QY 121 GTGCGGAGGGTGTGCGGCGGAGCTGGCCGGCTTGGGGGCGGAGTGGGATGTTGCG 180
DB 3724 CGAGCGGATGAGCTGGTGGCGAGCTGGAGCAGCGCGCGCGACGTGCGGTGCTCG 3783
QY 181 TGTGATGTGGGGAGCGCGGAGGTGTGCTCGGCTGCTGGAGGGTGTTCCTCGCGGTGT 240
DB 3784 TCGGACTCGGCAGATCGGGAATCGCTTGGCGCGCTGGTGGCGCTGCTCGGAAAC 3843
QY 241 CCGCTGACGGGTGTGTCGATCGGCTGGTGTGTCGACGATGCGACGATCGCTCTCTC 300
DB 3844 CCGTTGGGGTGGTGTGTCACCGCGCGGTGTGCTGGATGACGGTGTGTCGATGTC 3903
QY 301 AGCCCCGAGCGGTGGGCAAGTGTTCGCGGCCAAGTGGATGCGCTCTTTTGTGGAT 360
DB 3904 TCGCCGAGCGCTTGGACGCGGTGTTGCGGCCCAAGTGGATGCGCGGTGGTACCTGC 3963
QY 361 GAGCTACCGGGGTATGAGCTGTGCGGTTGCTGCTGTTCTCTCGCGCGCGGGATC 420
DB 3964 GAGCTGACTCGGGAATCGGCTGTGCGGCTTGTGTTGTTCTCTCGTGGCGGCTG 4023
QY 421 CTGGGGTGGCGCGGCGAGGCAATACGCGCGGCCAATGCGCTCTGGACGCGCTGGCG 480
DB 4024 TTCGGCGGTGGCGGGGAGAGCAATACGCTGCGCGCAACGCTTCTGGATGCTTGGCG 4083
QY 481 TACCGCGCGGCGCGGGTCTGCGCGGGGTGTGCTGCGGGGGCTGTGG 534
DB 4084 CATTGCGCGAGGCCCGCGGCTGCGCGCTGCTGCTGCGCTCGCGGCTGTGG 4137

CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=629432881.
XX
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 46.7%; Score 249.2; DB 22; Length 4403765;
Best Local Similarity 66.7%; Pred. No. 1.2e-37;
Matches 356; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1 GCGTCGGTCTGGTACCGGTCGGGACGGGTGCTGGGTGCGCGGTGCGCGGATCTG 60
DB 2298157 GGCACGGTGTGATACCGGTCGACCGGGATGGCGGGTTCGCGGTGCGCGTACG 2298098

QY 61 GCTGCTGTGTGGGTGCGGATCTGCTGTTGGTGAGCGCGGTGCTCCGATGCTCCG 120
DB 2298097 GTGGCTCGTCATGGGTGCGCAATCTGTTGGTGAGCGCGGCGCGGATGCTCC 2298038

QY 121 GGTGCGAGGCTGTCGGCGGAGCTGCGCGGTTCGGCGGAGGTGCGGATGTTGCG 180
DB 2298037 GGGGCTGCGGAGCTGCTGCGGAGTTGGCGCGGTTCGCGGATGAGGTGTCG 2297978

QY 181 TGTGATGTGGGAGCGCGGAGGTGTCGGCTGCTGGAGGTGTTCTCCCGGTCT 240
DB 2297977 TGTGACCGCGGATCGAGCGGTTGGCGGATGATCGCGATATCCGGTTACGAT 2297918

QY 241 CCGCTGACGGGTCTCGTGATCGGCTGCTGCTGGACGATCGACGATCGCTCTCTC 300
DB 2297917 CCATGTGCGGCTGATCACACCGCGCGGCTGCGAGCGGTGATGTCACATG 2297858

QY 301 AGCCCGAGCGCTGGGACGCTGTTGCGGCGCAAGTGGATGCGCTCTTTGCTGAT 360
DB 2297857 ACACCGGATCGGTGATGCTGTTGCGGCTCCAGGTGAGCGCGGTGCGACCTG 2297798

QY 361 GAGTCACCGGATGAGCTGCTGGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 2297797 GAGTTGACTGCGACCTGATGCTGCGGTTTGTGATGTTGCTGATGCGCGCTG 2297738

QY 421 CTGGGCTGCGCGGCGGCAACTACGCGCGGCAATGCGCTCTGACGCGCTGCGG 480
DB 2297737 GTCCGATGCTGCGGCGGCAACTATGCGCGCGCAATTCGTTTGGATGCGCTG 2297678

QY 481 TACCGCGGCGGCGCGGCTGTCGGGCGGTGCTGCTGCGGTGCGGCTGCGG 534
DB 2297677 GCCACCGGCGGCGGCTGCGGCTGCGGCGCATCTCCCTGCGGCTGCTG 2297624

RESULT 13
AAI17185
ID AAI17185 standard; DNA; 27541 BP.
XX
AC AAI17185;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
DE Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nys2; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (454..1191)
FT /*tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT complement (1275..3092)
FT CDS
FT /*tag= b

FT CDS
FT /*product= "NysG protein"
FT complement (3070..4824)
FT /*tag= c
FT /note= "CDS does not include start codon"
FT /*product= "NysH protein"
FT 5122..6156
FT /*tag= d
FT /*product= "NysD3 protein"
FT 6338..27541
FT /*tag= e
FT /*product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX
PN WO200159126-A2.
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UVNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingren TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
XX
P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2; Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

Query Match 45.7%; Score 244; DB 22; Length 27541;
Best Local Similarity 67.8%; Pred. No. 1.2e-36;
Matches 362; Conservative 0; Mismatches 160; Indels 12; Gaps 1;

QY 1 GGGTCGGTGTGCTGACGGGTGGGACGGGTGCTGGGTGCGCGGTGCGCGGATCTG 60
DB 10664 GGTGCGGTCCTGGTCACTGGTGGTACGGGTGGTCTGGGTGCGGTGCTGGCGGT 10723

QY 61 GCTGCTGTGTGTTGGGTGCGGATCTGCTGTTGGTGAGCGCGGTGCTCCGATGCTCCG 120
DB 10724 GTGGCCGAGTATGGGTGCGGATCTGCTGTTGGTCACTGCGAGTGGTACGCGGTG 10783

QY 121 GGTGCGAGGTCTGCGGCGGAGCTGCGCGGTGTTGGGGGCGGAGGTGCGGATTTGCG 180
DB 10784 GGTGCTGGGAGTTGGTGGCGAGCTTGGCGGTGTTGGGTGCGCGGTGCGGTTGCG 10843

QY 181 TGTGATGTGGGGAGCGGAGGTGCTCGGCTGCTGGAGGTGTTCTCTGCCGGGTCT 240
DB 10844 TGTGATGTGACCGATCGTGGCGCGGTGGTGGAGTTGGT-----TGGCGGCGAT 10891

Db 13224 GAACTGACCGGACGACCGCTGTGCGGTCTTCTCGCGGCTGCGCGCATC 13283
QY 421 CTGGGCTCGCCCGGCGAGGCAACTAGCGCGCGCCAAATCCGCTCTGGACGCGCTGGCG 480
Db 13284 CTGGCGCGCGCGGACAGGCAACTAGCGCGCGCCAAACACCTTCTCTGACGCGCTCGCG 13343
QY 481 TACCGCGCGCGCGCGGCTCTGCGCGCGGCTGCTGCGCTGCGCGGCTGCTGCG 534
Db 13344 CAGCACCGCGCGCGCCACAGCGCTCCCGCGGTGCTGCGCTGCGCGGCTGCTG 13397

RESULT 11
AA199682/c
ID AA199682 standard; DNA; 4411529 BP.
XX
AC AA199682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 47.0%; Score 250.8; DB 22; Length 4411529;
Best Local Similarity 66.9%; Pred. No. 6.1e-38;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 1 GGGTCGCTGTGTCACGGTGGGACGCGGTGCTGGGTGGCGCGGTGGCGGCACTCG 60
Db 2295857 GGCACGGTGTGATCACCGTGGCACCGGATGGCGGTTCGGCGGTGGCGGTCACGTG 2295798
QY 61 GCTGCTGTGTGGGTGGCGGATCTGCTGTTGGTGGACCGCGGTGGTCCGATCTCG 120

Db 2295797 GTGGCTGTCTATGGGTGGCGCAATCTGTGTTGGTAGCCCGCGCGGCGGATGCTCCC 2295738
QY 121 GGTGGCGAGGGTCTCGCGGCGGAGCTGGCGCGCTTGGGGCGGAGGTGCGGATGTTGTCG 180
Db 2295737 GGGGCTCGGAGCTGGTGGCCGAGTTGGCCGCGCGGCTGCCAGGTGCAAGTGTGGCC 2295678
QY 181 TGTGATGTGGGGAGCGCGGAGGTGTTCCGGCTGTGAGAGGTGTTCTCTGCCCGGTTGT 240
Db 2295677 TGTGACGCGCGGATCGAGCGCGCTTGGCCAAGGTGATCGCGATATTCCGGTTCAGCAT 2295618
QY 241 CCGCTGACCGGTGTCTGTCATCGCGCTGGTGTGCTGGACGATCGCGATCGCTCTCTC 300
Db 2295617 CCAATTGTCTGGCGTGTATCCACACCGCGCGCACTCGACGACGCTGGTGTGATGTCATG 2295558
QY 301 ACGCCGAGCGCTGCGGACGCGTGTTCGCGGCAAGGTGGATGCGCGCTCTTTTGTCTGAT 360
Db 2295557 ACACCGGATCGGTGTGATGTGTTGGTCCAAGGTGGACCGCGGTGGACCTTGCAC 2295498
QY 361 GAGCTGACCGCGGTATGAGAGCTGTGCGGTTCGTGCTGTTCTCTCGCGCGCGGATC 420
Db 2295497 GAGTTGACTCGCGACCTGATGTGTCGGCTTTTGTGTCATGTTTTCGTGATGCGCGCTG 2295438
QY 421 CTGGGCTCGCGCGGCGAGGCAACTACGCGCGCGCAATGCGCTCTGAGACGCGCTGGCG 480
Db 2295437 GTCGGATCTGTCGGGCGAGGCCAACTATGCGCGCGCAATTCGTTTTCGATCGCTGGCC 2295378
QY 481 TACCGCGCGCGCGCGGCTGTCGCGGGGTGTCGCTGGCGGTGGCGGCTGTTGG 534
Db 2295377 GCCCACCGCGCGCGCATGGCTGCCGCCATCTCCCTGGGCTGGGCTGTGG 2295324

RESULT 12
AA199683/c
ID AA199683 standard; DNA; 4403765 BP.
XX
AC AA199683;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 47.0%; Score 250.8; DB 22; Length 4411529;
Best Local Similarity 66.9%; Pred. No. 6.1e-38;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
Query Match 48.2%; Score 257.2; DB 22; Length 125401;
Best Local Similarity 67.6%; Pred. No. 4.4e-39; Mismatches 0; Gaps 0;
Matches 361; Conservative 0; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGGTGAACGGGTGGACGGGTGTGCTGGGTGCGCGGTGGCGGCATCTG 60
DB 86180 GGAACGGTGTGGTGAACGGGTGGTACGGGTGTGCTGGGTGGGTGGTGGCGCATCTG 86239
QY 61 GCTGGTGTGTGGGTGGGATCTGCTGTGGTGAACGGCGGTGGTGGATGCTCG 120
DB 86240 GTCCGGTGTGTGGGTGGGATCTGCTGTGGTGAACGGTGGTGGTGGGTGGTGG 86299
QY 121 GGTGGGAGGGTCTGGCGCGGAGCTGGCGGTGGCGGTGGCGGAGGTGGGATTTGGG 180
DB 86300 GGTGGCGCGGTGGTGGCGGAGTTGGAGTCTGGCGCGCGGTGGTGGTGGCGG 86359
QY 181 TGTGATGTGGGAGCGCGGAGGTGGTCCGCTCTGGAGGGTGTCTCTGCGGGTGT 240
DB 86360 TGTGATGTGGTGAATGCTCGCTGTGGAGTGGTGGCGGTGGTGGAGTGGTAT 86419
QY 241 CGCTGACGGGTGTGTCGATGCGGTGTGGTGTGGAGTGGAGTGGAGTGGAGTGGAT 300
DB 86420 CGTGTCTCGGTGTGTCGATGCGGTGTGGTGTGGATGACGGTGTGGTGGTGGTGG 86479
QY 301 AGCCCGAGCGGTGGGACGGTGGTGGCGGCAAGGTGGATGCCCTTTTGTGCGAT 360
DB 86480 AGCCCGAGCGGTGGTGGCGGTGGTGGCGGCAAGGTGGATGGTGGTGGATGGAT 86539
QY 361 GAGCTACCGGGGTATGAGCTGTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420
DB 86540 GAGCGACGGTGTGGATCTGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 86599
QY 421 CTGGGTGGCGCGGAGGCAACTAGCGCGCGGCAATGCCGCTCTGACGGCTGGCG 480
DB 86600 TTGGGGGTGGCGGTTCAGGCAACTATGGCGGGTAAATGCGTTTGGACGGTGGATG 86659
QY 481 TACCGCGCGCGCGGTGTGGCGGGGTGTGGCGGGTGTGGCGGGGTGTGGCGGGTGTGG 534
DB 86660 GTTCATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 86713
RESULT 10
AAT80413
ID AAT80413 standard; DNA; 43280 BP.
AC AAT80413;
XX
XX
XX 27-FEB-1998 (first entry)
DE Tyalactone synthase gene cluster.
XX Tyalactone synthase gene cluster; tylG gene; multifunctional protein;
KW polyketide; tyalactone synthesis; antibiotic; tylosin; ss.
XX
XX Streptomyces fradiae.
XX
XX Key Location/Qualifiers
PH 816..14243
FT CDS
FT /tag= a
FT /transl_except= (pos: 816..818, aa: Met)
FT /note= "ORF1 encodes protein shown in AAW22601"
FT CDS
FT 14351..19945
FT /tag= b
FT /transl_except= (pos: 14351..14353, aa: Met)
FT /note= "ORF2 encodes protein shown in AAW22602"
FT

FT CDS 20010..31199
FT /tag= c
FT /transl_except= (pos: 20010..20012, aa: Met)
FT /note= "ORF3 encodes protein shown in AAW22603"
FT CDS 31232..36067
FT /tag= d
FT /note= "ORF4 encodes protein shown in AAW22604"
FT CDS 36249..41774
FT /tag= e
FT /note= "ORF5 encodes protein shown in AAW22605"
XX
XX EP791655-A2.
XX 27-AUG-1997.
XX 19-FEB-1997; 97EP-0301056.
XX 22-FEB-1996; 96US-0012078.
XX (ELIL) LILLY & CO ELI.
XX Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;
XX WPI; 1997-418046/39.
XX P-P8DB; AAW22601-W22605.
XX DNA encoding Streptomyces fradiae tyalactone synthase domain - for
XX production of tylosin-related polyketide compounds
XX Claim 2; Pages 8-66; 220pp; English.
XX This sequence represents the tyalactone synthase gene cluster of the
XX invention. This sequence is also referred to as the tylG gene, and was
XX isolated from Streptomyces fradiae. This sequence encodes multifunctional
XX proteins which direct the synthesis of the polyketide tyalactone, isolated
XX from Streptomyces fradiae. Tyalactone is the basic building block of the
XX antibiotic tylosin. The DNA sequence can be modified so as to alter the
XX type of carboxylic acids incorporated, the number of carboxylic acids
XX incorporated and/or the post-condensation reactions performed, thereby
XX resulting in novel tylosin-related polyketides.
XX
SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;
Query Match 47.0%; Score 250.8; DB 18; Length 43280;
Best Local Similarity 66.9%; Pred. No. 6.9e-38;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGGTGAACGGGTGGACGGGTGTGCTGGGTGGCGGTGGCGGCATCTG 60
DB 12864 GGCACGGTCTGGTGAACGGGTGGACGGGTGTGCTGGGTGGCGGTGGCGGCATCTG 12923
QY 61 GCTGGTGTGTGGGTGGGATCTGCTGTGGTGAACGGGTGGTGGTGGTGGTGGTGGTGG 120
DB 12924 GCCCGGAGACGGGTGTGGCGCATCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 12983
QY 121 GGTGGGAGGGTCTGGCGCGGAGCTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCG 180
DB 12984 GGGCGGAGCACTGCGCGGAACTGGCGGGTTGGAAGCCGCGGTGTGCTGGCGGC 13043
QY 181 TGTGATGTGGGAGCGCGGAGGTGGTTCGGGTGTGCTGGAGGTGTGCTTCGCGGGTGT 240
DB 13044 TGTGACGGCGGACCGGAGCCCTGGCAACGGGTGTGCTGGCGGTGGCGGTGGCGGC 13103
QY 241 CGCTGAGGGGTGTGCTGATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
DB 13104 CCGCTGACGGGAGTGAATCAGCGCGCGGGGTGTGCTGACGCGGACACTGACCGGCTG 13163
QY 301 AGCCCGGAGCGGTGGGACCGGTGTTCGCGGCAAGGTGGATGCGGCTCTTTTGTGGAT 360
DB 13164 ACCCGGAGCACTGCTGATCGGTACCGTCAATCGCGCGGAGCGCGGTGAACTGCAC 13223
QY 361 GAGCTGACCGGGGTATGAGGTGTGCGGTGTGCTGGTGTGCTGTGCTGTGCTGTGCTGTG 420

```
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
PI
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
PT
XX
XX Claim 2; Page 116-151; 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nysI DNA of nystatin PKS gene cluster.
XX
XX Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
SQ
Query Match 48.2%; Score 257.2; DB 22; Length 65140;
Best Local Similarity 67.6%; Pred. No. 4.5e-39;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;
QY 1 GCGTCGGTGTGTGGTACGGGTGGAGCGGTGTCTGGGTGGCGGTGGCGGATCTCG 60
Db |||||
QY 25919 GGAACGGTGTGTGGTACGGGTGTGTGGTGGTGTGGTGGTGGTGGTGGTGGT 25978
Db |||||
QY 61 GCTGGTGTGTGGTGGTGGGATCTGTCTGGTGGTGGTGGTGGTGGTGGTGGTGG 120
Db |||||
QY 25979 GTGCGGTGTGTGGTGGTGGGATCTGTCTGGTGGTGGTGGTGGTGGTGGTGG 26038
Db |||||
QY 121 GGTGCGGAGGTCTGCGGGGAGTGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGG 180
Db |||||
QY 26039 GGTGCGGCGGGTGTGTCGCGAGTGGAGTGTGTCGCGGCGCGGGTGTGTCG 26098
Db |||||
QY 181 TGTGATGTGGGAGCGCGGAGGTGTCTCGGCTGTGGAGGTGTCTCGCGGTGT 240
Db |||||
QY 26099 TGTGATGTGGGAGTGTGTCGCGGTGTGGAGGTGTGTCGCGGTGTGTCGAG 26158
Db |||||
QY 241 CCGCTGACGGGTGTCTGTCATGCGGTGTGTGGTGGTGGTGGTGGTGGTGGTGGT 300
Db |||||
QY 26159 CCGTGTCTGCGGTGTGTCATGCGGTGTGTGGTGGTGGTGGTGGTGGTGGTGG 26218
Db |||||
QY 301 AGCCCGAGCGGTGGGACGGTGTGCGGGGCAAGTGGATGCGGCTCTTTTGTCTGG 360
Db |||||
QY 26219 AGCCCGAGCGGTGGTGGTGGTGTGTCGCGGTGGTGGTGGTGGTGGTGGTGG 26278
Db |||||
QY 361 GAGCTGACCGGGGTATGAGTGTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTG 420
Db |||||
QY 26279 GAGGCGACCGGTGTGTCGATGCGGCTGTGGACGGTGTGTCGCTGCTGCTG 26338
Db |||||
QY 421 CTGGGGTCTGGCGGGGAGGCGCAACTACGCGCGGGGCAATGCGGCTGTGGACG 480
Db |||||
QY 26339 TTGCGGGGTGCGGGTACAGCCCACTATGCGCGGGGTAAATGCGTTTGGACG 26398
Db |||||
QY 481 TACCGCGCGCGCGCGGCTGTGCGGGGGTGTGCTGCGGGGGTGTGCGGGGGTGT 534
Db |||||
QY 26399 GTTCATCGGGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 26452
Db |||||
RESULT 9
AAD17186
ID AAD17186 standard; DNA; 125401 BP.
```

```
XX
AC AAD17186;
XX
XX 29-NOV-2001 (first entry)
XX
XX Streptomyces noursei nystatin PKS gene cluster DNA.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; ds.
XX
XX Streptomyces noursei.
XX
XX Key Location/Qualifiers
XX CDS 6337..34771
XX FT /*tag= a
XX FT /product= "NysI complete protein"
XX CDS 34792..51099
XX FT /*tag= b
XX FT /product= "NysJ protein"
XX CDS 51155..57355
XX FT /*tag= c
XX FT /product= "NysK protein"
XX CDS 57503..58687
XX FT /*tag= d
XX FT /product= "NysL protein"
XX CDS complement (58786..58980)
XX FT /*tag= e
XX FT /product= "NysM protein"
XX FT /note= "CDS does not include start codon"
XX CDS complement (59045..60241)
XX FT /*tag= f
XX FT /product= "NysN protein"
XX FT /note= "CDS does not include start codon"
XX CDS complement (60238..61296)
XX FT /*tag= g
XX FT /product= "NysD2 complete protein"
XX CDS 120628..121308
XX FT /*tag= h
XX FT /product= "NysR4 (long) protein"
XX
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX 10-APR-2000; 2000GB-0008786.
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
XX (ALPH-) ALPHARMA AS.
XX (SINV-) SINVENT AS.
XX (DZIE/) DZIEGLEWSKA H.
XX (ZOTC/) ZOTCHEV S B.
XX (SEKU/) SEKUROVA O N.
XX (FJAE/) FJAERVIK E.
XX (BRAU/) BRAUTASET T.
XX (STRO/) STROM A R.
XX
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
PI
XX
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
XX Claim 1; Page 188-254; 266pp; English.
XX
```

XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARNA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TB, Sletta H, Gulliksen O;
XX WPI; 2001-557614/62.
DR P-PSDB; AAEL10153.
XX New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals. -
XX Example 2; Page 186-187; 266pp; English.
XX The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a ERD48 insert DNA in pKR48 which is used in an
CC exemplification of the invention.
XX
XX Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;
SQ

Query Match 48.2%; Score 257.2; DB 22; Length 2700;
Best Local Similarity 67.6%; Pred. No. 4.9e-39;
Matches 361; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1 GGTCGGTGTGTGTCACGGGTGGGACGGGTGTCTCGGTCGGCGGTGGCGGCATCTG 60
DB |||||
QY 61 GCTGGTGTGTGCGGTGGGATCTGCTTGTGTCAGCGCGGTGTCGGATGTCG 120
DB |||||
QY 1230 GTCCGGTGTGTGCGGTGGGATCTGCTTGTGTCAGCGGTGTCGGATGTCG 1289
DB |||||
QY 121 GTGCGGAGGCTCTCGCGCGGAGCTGGCGCTTGGCGCGGAGTGGATGTCG 180
DB |||||
QY 1290 GTGCGCGCGGTGTGTCGGGAGTTGGAGTCTGTGGCGCGCGGTGTGTCG 1349
DB |||||
QY 181 TGTGATGTGGGGAGCGCGGAGGTGGTTCGGCTCTCGAGGGTGTCTCTCGCGGTGT 240
DB |||||
QY 1350 TGTGATGTGGGTATGCTCGGCTGTGCGAGTGTGTTGCGGTGTGTCGAGTCTAT 1409
DB |||||
QY 241 CCGCTACGGGTGTGTCATGCGGTGTGTCGTCGACGATCGGATCTCTCTC 300
DB |||||
QY 1410 CCGTGTCTGCGGTGTGTCATGCGGTGTGTCGATGACGATGTCGTTGCGGT 1469
DB |||||
QY 301 ACCCGCGAGCGGTGGCGAGCTGTTCGCGGCAAGTGGATCCGCTCTTTTGTGGAT 360
DB |||||
QY 1470 ACCCGCGAGCGGTGGCGAGTGTGTCGTCGCAAGTGGATGTCGTTGCGGTG 1529
DB |||||
QY 361 GAGCTACCGCGGTATGAGAGTGTTCGGCTTCGTCGTCCTCTCGCGCGCGGATC 420
DB |||||
QY 1530 GAGCGACGCGGTCTGGATCTGGACGCGTGTGTTGTCCTCTCTGTCGTCG 1589
DB |||||
QY 421 CTGGGTGCGCGCGGAGGCACTACGCGCGGCAATGCGCTCTGACGCGCTGGCG 480
DB |||||
QY 1590 TTCGGGGTTCGGGTTCAGGCCAACTATGCGCGGGTAATGCGTTTGGACGCGTGTG 1649
DB |||||
QY 481 TACCGCGCGCGGTGTGTCGCGGGGTGTGTCGTCGCGGTGTCGTCGTCGTCG 534
DB |||||
QY 1650 GTTCATCGGTGCTGTGTGGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1703
DB |||||

RESULT 8
AADI7184
ID AADI7184 standard; DNA; 65140 BP.
XX
AC AADI7184;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; nys1; ds.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS complement (1..1035)
FT /*tag= a
FT /product= "NysD2 partial protein"
FT /note= "CDS does not include stop codon"
FT CDS complement (1056..2576)
FT /*tag= b
FT /product= "NysD1 protein"
FT 2806..6906
FT /*tag= c
FT /product= "NysA protein"
FT 6952..16530
FT /*tag= d
FT /product= "NysB protein"
FT 16550..49840
FT /*tag= e
FT /product= "NysC protein"
FT 50260..51015
FT /*tag= f
FT /product= "NysE protein"
FT 51405..54305
FT /*tag= g
FT /product= "NysR1 protein"
FT 54329..57190
FT /*tag= h
FT /product= "NysR2 protein"
FT /note= "CDS does not include start codon"
FT 57180..59963
FT /*tag= i
FT /product= "NysR3 protein"
FT 60415..61047
FT /*tag= j
FT /product= "NysR4 (short) protein"
FT /note= "CDS does not include start codon"
FT 61736..62497
FT /*tag= k
FT /product= "NysR5 protein"
FT /note= "CDS does not include start codon"
FT complement (62551..63615)
FT /*tag= l
FT /product= "ORF2 protein"
FT /note= "CDS does not include start codon"
FT 63765..64961
FT /*tag= m
FT /product= "ORF1 protein"
XX
XX WO200159126-A2.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
XX
XX 10-APR-2000; 2000GB-0008786.
XX
XX 14-APR-2000; 2000GB-0009387.
XX
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.

Db 19525 CCGCTGACGGGTGCTGTCATCGCGTGTGTGCTGGACGATGCGATCGCCTCTCTC 19584
 QY 301 ACSCCCGAGCGCTGGCCACGGTGTTCGCGGCCAAGTGGATGCCCGCTCTTTTGTGGAT 360
 Db 19585 ACSCCCGAGCGCTGGCCACGGTGTTCGCGGCCAAGTGGATGCCCGCTCTTTTGTGGAT 19644
 QY 361 GAGCTGACCGGGGATGAGAGCTGTTCGGCGTTCTGCTGTCTCTCGCCCGCGGGATC 420
 Db 19645 GAGCTGACCGGGGATGAGAGCTGTTCGGCGTTCTGCTGTCTCTCGCCCGCGGGATC 19704
 QY 421 CTGGGGTCCGCGCGGACAGGCACTACGCGCGGCGCAATGCCGCTCTGACCGGCTGGCG 480
 Db 19705 CTGGGGTCCGCGCGGACAGGCACTACGCGCGGCGCAATGCCGCTCTGACCGGCTGGCG 19764
 QY 481 TACCGCGCGCGGCGCGGCTGTGCGCGGGGTGTCTGCGCGGGCGGTGTGG 534
 Db 19765 TACCGCGCGCGGCGCGGCTGTGCGCGGGGTGTCTGCGCGGGGTGTGG 19818

RESULT 6

AAH79278

ID AAH79278 standard; DNA; 31422 BP.

XX

AC AAH79278;

DT 04-DEC-2001 (first entry)

XX

DE Streptomyces avermitilis coding sequences SEQ ID NO: 2.

XX

KW Avermectin aglycone synthase; AAS; avermectin derivative;
 drug production; veterinary drug; pesticide; ds.

XX

OS Streptomyces avermitilis.

XX

FH Key Location/Qualifiers

FT 1..14646

FT /*tag= a

FT /product= "AAG65266"

FT 14824..31422

FT /*tag= b

FT /product= "AAG65267"

XX WO200162939-A1.

XX

XX 30-AUG-2001.

XX

XX 23-FEB-2001; 2001WO-JP01381.

XX

XX 24-FEB-2000; 2000JP-0047405.

XX

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX

XX (KITA) KITASATO INST.

XX

XX Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;

XX

XX WPI; 2001-582053/65.

XX

XX P-PSDB; AAG65266, AAG65267.

XX

XX New modified avermectin aglycone synthase derived from Streptomyces

XX

XX avermectilis used in production of 22,23-dihydroavermectin B1a used in

XX

XX drugs and pesticides -

XX

XX Disclosure; Page 103-149; 257pp; Japanese.

XX

CC The present invention relates to the production of modified derivatives
 of avermectin aglycone synthase (AAS) derived from Streptomyces
 avermitilis. The activity of an acyl carrier protein (ACP),
 beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 reductase (ER) and/or thioesterase (TE) domain may be reduced or
 suppressed. The process can be used in the production of drugs, veterinary
 drugs and pesticides. The present sequence is a fragment of the S.
 avermitilis genome.

XX SQ Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;
 Query Match 99.7%; Score 532.4; DB 22; Length 31422;
 Best Local Similarity 99.8%; Pred. No. 8.3e-90;
 Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GGGTCGGTGTGGTGACCGGGTGGGACGGGTGTGCTGGGTGCGCGCGGTGGCGCGCATCTG 60
 Db 19285 GGGTCGGTGTGGTGACCGGGTGGGACGGGTGTGCTGGGTGCGCGCGGTGGCGCGCATCTG 19344
 QY 61 GCTGTGTGTGTGGGTGCGGATCTCTGTGTGTGTGAGCCGCGTGTCTCCGATGCTCCG 120
 Db 19345 GCTGTGTGTGTGGGTGCGGATCTCTGTGTGTGTGAGCCGCGTGTCTCCGATGCTCCG 19404
 QY 121 GGTTCGGAGGGTCTCGCGGCGGAGCTGCGCGGCTGGGGCGGAGTGGGATTTGTTGG 180
 Db 19405 GGTTCGGAGGGTCTCGCGGCGGAGCTGCGCGGCTGGGGCGGAGTGGGATTTGTTGG 19464
 QY 181 TGTGATGTGGGGAGCGCGGAGGTGTGTCGGCTGTCTGGAGGGTGTTCCTGCCGGGTGT 240
 Db 19465 TGTGATGTGGGGAGCGCGGAGGTGTGTCGGCTGTCTGGAGGGTGTTCCTGCCGGGTGT 19524
 QY 241 CCGCTGACCGGTGTCTGTCATCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
 Db 19525 CCGCTGACCGGTGTCTGTCATCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19584
 QY 301 ACGCCCGAGCGCTGGGACCGGTGTTCGGCGCCAAAGTGGATGCCCGCTCTTTTGTGGAT 360
 Db 19585 ACGCCCGAGCGCTGGGACCGGTGTTCGGCGCCAAAGTGGATGCCCGCTCTTTTGTGGAT 19644
 QY 361 GAGCTGACCGGGGTATGAGAGCTGTGCGGCTGTGCTGCTGCTGCGCGCGGGGATC 420
 Db 19645 GAGCTGACCGGGGTATGAGAGCTGTGCGGCTGTGCTGCTGCTGCGCGCGGGGATC 19704
 QY 421 CTGGGGTCCGCGCGGCGAGGCAACTACCGCGCGCAATGCGCTCTGTGACCGCGTGGCG 480
 Db 19705 CTGGGGTCCGCGCGGCGAGGCAACTACCGCGCGCAATGCGCTCTGTGACCGCGTGGCG 19764
 QY 481 TACCGCGCGCGGCGCGGCTGTGCGCGGGGTGTGCGCGGGGTGTGCGCGGGGTGTGG 534
 Db 19765 TACCGCGCGCGGCGCGGCTGTGCGCGGGGTGTGCGCGGGGTGTGG 19818

RESULT 7
 AAD17193
 ID AAD17193 standard; DNA; 2700 BP.
 XX
 AC AAD17193;
 XX
 DT 29-NOV-2001 (first entry)
 XX
 DE ERD48 insert DNA in pXR48.
 XX
 KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
 antifungal; antibiotic; ds.
 XX
 OS Streptomyces noursei.
 XX
 FH Key Location/Qualifiers
 CDS 3..2699
 FT /*tag= a
 FT /product= "Protein encoded by ERD48 insert DNA"
 XX
 PN WO200159126-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 08-FEB-2001; 2001WO-GB00509.
 XX
 PR 08-FEB-2000; 2000GB-0002840.
 PR 10-APR-2000; 2000GB-0008786.
 PR 14-APR-2000; 2000GB-0009387.

XX PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX WPI; 2001-582053/65.
DR P-PSDB; AAG65264, AAG65265.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Example 2; Page 58-123; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermectilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, veterinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermectilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
Query Match 100.0%; Score 534; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred. No. 4.2e-90;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGGTACGGGTGGGACGGGTGTGCTGGGTGCGCGGTGCGCGGATCTG 60
Db 10609 GGGTCGGTGTGGTACGGGTGGGACGGGTGTGCTGGGTGCGCGGTGCGCGGATCTG 10668
QY 61 GCTGGTGTGCTGGGTGGGATCTGCTGGTGGAGCGCGGTGCTCGGATCTCCG 120
Db 10669 GCTGGTGTGCTGGGTGGGATCTGCTGGTGGAGCGCGGTGCTCGGATCTCCG 10728
QY 121 GGTGGGAGGCTCTGGCGGCGAGCTGGCGGCTTGGGGCGGAGGTGGGATGTTGG 180
Db 10729 GGTGGGAGGCTCTGGCGGCGAGCTGGCGGCTTGGGGCGGAGGTGGGATGTTGG 10788
QY 181 TGTGATGTGGGAGCGCGGAGGTGTCTCGGCTCTGGAGGTGTTCTTCGCGGGTGT 240
Db 10789 TGTGATGTGGGAGCGCGGAGGTGTCTCGGCTCTGGAGGTGTTCTTCGCGGGTGT 10848
QY 241 CGCTACGGGTGTCTGTCATCGGTGGTGTGCTGGAGATGCGACGATCGCTCTTC 300
Db 10849 CGCTACGGGTGTCTGTCATCGGTGGTGTGCTGGAGATGCGACGATCGCTCTTC 10908
QY 301 ACGCCGAGCGGCTGGCGGCGAGCTGTTCGCGGCGCAAGTGGATGCGCTCTTTTGTGGAT 360
Db 10909 ACGCCGAGCGGCTGGCGGCGAGCTGTTCGCGGCGCAAGTGGATGCGCTCTTTTGTGGAT 10968
QY 361 GAGCTACCGGGGTATGAGCTGTGCGGCTTCTGCTGTCTTCCTTCGCGCGCGGGATC 420
Db 10969 GAGCTACCGGGGTATGAGCTGTGCGGCTTCTGCTGTCTTCCTTCGCGCGCGGGATC 11028
QY 421 CTGGGGTTCGCGCGGCGAGGCACTACGCGCGGCGCAATGCGCTCTGAGCGCGTGGCG 480
Db 11029 CTGGGGTTCGCGCGGCGAGGCACTACGCGCGGCGCAATGCGCTCTGAGCGCGTGGCG 11088
QY 481 TACCGCGCGGCGCGGCTCTGCGCGGGGTGTCTGCTGGCGGGGTGTGG 534
Db 11089 TACCGCGCGGCGCGGCTCTGCGCGGGGTGTCTGCTGGCGGGGTGTGG 11142
RESULT 5
ID AAA92302
XX AAA92302 standard; DNA; 31422 BP.
XX
AC AAA92302;
XX
XX 10-JAN-2001 (first entry)
DX

DE S. avermectilis avermectin aglycone synthase DNA aveAII SEQ ID NO:2.
XX
KW Streptomyces avermectilis; avermectin aglycone synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermectilis.
XX
FH Key Location/Qualifiers
FT CDS 1..14646
FT /*tag= a
FT /note= "avermectin aglycone synthase protein"
FT CDS 14824..31422
FT /*tag= b
FT /note= "avermectin aglycone synthase protein"
XX
PN WO200050605-A1.
XX
XX 31-AUG-2000.
XX 23-FEB-2000; 2000WO-JP01041.
XX 24-FEB-1999; 99JP-0046961.
XX (KITA) KITASATO INST.
XX Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR P-PSDB; AAB23751, AAB23752.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 2; Page 134-203; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
Query Match 99.7%; Score 532.4; DB 21; Length 31422;
Best Local Similarity 99.8%; Pred. No. 8.3e-90;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGTCGGTGTGGTACGGGTGGGACGGGTGTGCTGGGTGCGCGGTGCGCGGATCTG 60
Db 19295 GGGTCGGTGTGGTACGGGTGGGACGGGTGTGCTGGGTGCGCGGTGCGCGGATCTG 19344
QY 61 GCTGCTGTGTGGGTGCGGATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
Db 19345 GCTGCTGTGTGGGTGCGGATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19404
QY 121 GGTGGGAGGCTCTCGCGCGGAGCTGGCGGCTTGGGGCGGAGGTGGGATGTTGG 180
Db 19405 GGTGGGAGGCTTTCGCGGCGGAGCTGGCGGCTTGGGGCGGAGGTGGGATGTTGG 19464
QY 181 TGTGATGTGGGAGCGCGCGGAGGTGTGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
Db 19465 TGTGATGTGGGAGCGCGCGGAGGTGTGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 19524
QY 241 CCGCTGACGGGTGTCTGTCATCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300

Db 11786 GAGCTGACGGGGTATGAGCTGTGCGGCTTCGTCTCTCTCCCTCGGCGCGGGATC 11845
 QY 421 CTGGGGTCCGCGCGGAGGCAACTACGCGCGGCAATGCGCTCTGGACGCGCTGGCG 480
 Db 11846 CTGGGGTCCGCGCGGAGGCAACTACGCGCGGCAATGCGCTCTGGACGCGCTGGCG 11905
 QY 481 TACCGCGCGGCGCGGCTCTGCGCGGGGGTGTGCTGCGCTGGGGGCTGTGG 534
 Db 11906 TACCGCGCGGCGCGGCTCTGCGCGGGGGTGTGCTGCGCTGGGGGCTGTGG 11959

RESULT 3

AAA92301
 ID AAA92301 standard; DNA; 30690 BP.

AC AAA92301;

XX 10-JAN-2001 (first entry)

XX S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.

XX Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
 KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
 KW agrochemical; ds.

XX Streptomyces avermitilis.

XX Key Location/Qualifiers
 FH CDS 1..11919

FT /*tag= a

FT /note= "Avermectin aglycon synthase protein"

FT CDS 11971..30690

FT /*tag= b

FT /note= "Avermectin aglycon synthase protein"

XX WO200050605-A1.

XX 31-AUG-2000.

XX 23-FEB-2000; 2000WO-JP01041.

XX 24-FEB-1999; 99JP-0046961.

XX (KITA) KITASATO INST.

XX Omura S, Ikeda H;

XX WPI; 2000-565458/52.

XX P-PSDB; AAB23749, AAB23750.

XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
 PT it for the production of avermectin and its derivatives for drug and
 PT agrochemical use

XX Claim 2; Page 66-134; 314pp; Japanese.

XX The present sequence represents DNA which encodes avermectin aglycon
 CC synthase proteins. Also described are: (1) polypeptides encoded by all
 CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
 CC cells transformed by the vectors; (4) preparation of the polypeptides
 CC by culture of the transformants; (5) preparation of avermectin aglycon
 CC or its derivatives by culture of transformed avermectin-producing
 CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
 CC containing sense or antisense sequences from the avermectin aglycon
 CC synthase DNA. The enzymes are useful for the production of modified
 CC forms of avermectin and of the intermediates in its biosynthesis, for
 CC use as drugs, veterinary drugs and agrochemicals.

XX Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

XX Query Match 100.0%; Score 534; DB 21; Length 30690;

XX Best Local Similarity 100.0%; Pred. No. 4.2e-90;

Matches 534; Conservative 0; Mismatches 0; Indels. 0; Gaps 0;
 QY 1 GGGTCCGTGTGCTGACGGGTGGGACGGGTGTGCTGGGTGCGCGCGGATCTG 60
 Db 10609 GGGTCCGTGTGCTGACGGGTGGGACGGGTGTGCTGGGTGCGCGCGGATCTG 10669
 QY 61 GCTGTGTGTGTGGGTGCGGATCTGCTTTGTGTGAGCCGCGTGGTCCGATGCTCG 120
 Db 10669 GCTGTGTGTGTGGGTGCGGATCTGCTTTGTGTGAGCCGCGTGGTCCGATGCTCG 10728
 QY 121 GGTGCGAGAGGTCTCCGGGCGAGAGTGGCGGCTGGGGGCGAGGTGCGGATGTTGTGCG 180
 Db 10729 GGTGCGAGAGGTCTCCGGGCGAGAGTGGCGGCTGGGGGCGAGGTGCGGATGTTGTGCG 10788
 QY 181 TGTGATGTGGGGAGCGCGGAGGTGCTCGGCTGTGGAGGGGTTCCTCCCGGCTGT 240
 Db 10789 TGTGATGTGGGGAGCGCGGAGGTGCTCGGCTGTGGAGGGGTTCCTCCCGGCTGT 10848
 QY 241 CCGCTGACGGGTGTGCTGATCGGCTGGTGTGTGTGAGCGATGCGACGATCGCCTCTCTC 300
 Db 10849 CCGCTGACGGGTGTGCTGATCGGCTGGTGTGTGTGAGCGATGCGACGATCGCCTCTCTC 10908
 QY 301 AGCCCGGAGCGGCTGGGACAGGTGTTCCGGGCAAGGTGGATGCGCTCTTTTGTGTGAT 360
 Db 10909 AGCCCGGAGCGGCTGGGACAGGTGTTCCGGGCAAGGTGGATGCGCTCTTTTGTGTGAT 10968
 QY 361 GAGCTGACGCGGGGTATGAGCTGTCGCGTTCGTTGCTGCTCCGCGCGCGGGATC 420
 Db 10969 GAGCTGACGCGGGGTATGAGCTGTCGCGTTCGTTGCTGCTCCGCGCGCGGGATC 11028
 QY 421 CTGGGTGCGCGCGGAGGCAACTACGCCGCGGCAATGCGCTCTTGAGACGCGTGGCG 480
 Db 11029 CTGGGTGCGCGCGGAGGCAACTACGCCGCGGCAATGCGCTCTTGAGACGCGTGGCG 11088
 QY 481 TACCGCGCGGCGCGGCTGCGCGGGGTGCTGCGGGGGTGTGCTGGGGGCTGTGG 534
 Db 11089 TACCGCGCGGCGCGGCTGCGCGGGGTGCTGCGGGGGTGTGCTGGGGGCTGTGG 11142

RESULT 4

AAH79277

ID AAH79277 standard; DNA; 30690 BP.

XX AC AAH79277;

XX 04-DEC-2001 (first entry)

XX Streptomyces avermitilis coding sequences SEQ ID NO: 1.

XX Avermectin aglycone synthase; AAS; avermectin derivative;

XX drug production; veterinary drug; pesticide; ds.

XX Streptomyces avermitilis.

XX Key Location/Qualifiers
 FH CDS 1..11919

FT /*tag= a

FT /product= "AAG65264"

FT CDS 11971..30690

FT /*tag= b

FT /product= "AAG65265"

XX WO200162939-A1.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-JP01381.

XX 24-FEB-2000; 2000JP-0047405.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX (KITA) KITASATO INST.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
 XX WPI: 2001-582053/65.
 DR P-PSDB; AAG65268.
 XX
 XX New modified avermectin aglycone synthase derived from Streptomyces
 PT avermitilis used in production of 22,23-dihydroavermectin B1a used in
 PT drugs and pesticides -
 XX
 XX Claim 10; Page 149-167; 257pp; Japanese.
 XX
 XX The present invention relates to the production of modified derivatives
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces
 CC avermitilis. The activity of an acyl carrier protein (ACP),
 CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
 CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
 CC suppressed. The process can be used in the production of drugs, veterinary
 CC drugs and pesticides. The present sequence is modified version of a
 CC fragment of the S. avermitilis genome.
 XX
 SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;
 Query Match 100.0%; Score 534; DB 22; Length 11916;
 Best Local Similarity 100.0%; Pred. No. 4.3e-90;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGTCGGTGTGGTACGCGGTGGACGGGTGCTGGGTGGCGGTGGCGGCATCTG 60
 DB 10609 GGGTCGGTGTGGTACGCGGTGGACGGGTGCTGGGTGGCGGTGGCGGCATCTG 10668
 QY 61 GCTGGTGTGTGGGTGGCGATCTGCTGGTGGAGCCGGTGGTCCGATGCTCG 120
 DB 10669 GCTGGTGTGTGGGTGGCGATCTGCTGGTGGAGCCGGTGGTCCGATGCTCG 10728
 QY 121 GGTGGAGGGTGTGGCGCGAGTGTGCGCGTGGTGGGGCGGAGTGGGATTTGCG 180
 DB 10729 GGTGGAGGGTGTGGCGCGAGTGTGCGCGTGGTGGGGCGGAGTGGGATTTGCG 10788
 QY 181 TGTGATGTGGGGAGCGCGGAGGTGTCGGCTGTGGAGGTGTTCTCCCGGTCT 240
 DB 10789 TGTGATGTGGGGAGCGCGGAGGTGTCGGCTGTGGAGGTGTTCTCCCGGTCT 10848
 QY 241 CGGCTGACCGGGTGTCTGATCGCGTGGTGTGGAGTGGACGATCGCCTCTCTC 300
 DB 10849 CGGCTGACCGGGTGTCTGATCGCGTGGTGTGGAGTGGACGATCGCCTCTCTC 10908
 QY 301 ACGCCGAGCGGTGGGCAAGGTGTTCCGGCGCAAGTGGATGCCGCTCTTTTGTGGAT 360
 DB 10909 ACGCCGAGCGGTGGGCAAGGTGTTCCGGCGCAAGTGGATGCCGCTCTTTTGTGGAT 10968
 QY 361 GAGCTGACCGGGTATGAGCTGTGGGTTCGTGCTGTTCTCTCCGCGCGGGATC 420
 DB 10969 GAGCTGACCGGGTATGAGCTGTGGGTTCGTGCTGTTCTCTCCGCGCGGGATC 11028
 QY 421 CTGGGGTCCGCGGGGAGGCAACTACGCGGGGCAATGCGCTCTGACGCGTGGG 480
 DB 11029 CTGGGGTCCGCGGGGAGGCAACTACGCGGGGCAATGCGCTCTGACGCGTGGG 11088
 QY 481 TACCGCGCGGCGCGGTCTGCGGGGGGTGTCGTCGCGGTGGGGGTGTGG 534
 DB 11089 TACCGCGCGGCGCGGTCTGCGGGGGGTGTCGTCGCGGTGGGGGTGTGG 11142
 RESULT 2
 AAZ58381
 ID AAZ58381 standard; DNA; 12381 BP.
 XX
 XX AAZ58381;
 AC
 XX
 XX 23-MAY-2000 (first entry)
 DT
 XX Streptomyces avermitilis avermectin polyketide synthase modules 1+2.

XX Polyketide synthase; avermectin; insecticide; ss.
 KW Streptomyces avermitilis.
 OS
 XX WO200001827-A2.
 PN
 XX 13-JAN-2000.
 PD
 XX 06-JUL-1999; 99WO-GB02158.
 PF
 XX 06-JUL-1998; 98GB-0014622.
 PR
 XX (BIOTIC) BIOTICA TECHNOLOGY LTD.
 PA (PTIZ) PFIZER INC.
 XX Kallenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
 PI McArthur HAI;
 XX WPI: 2000-182117/16.
 DR
 XX Mutated Type I polyketide synthase containing a polylinker site in an
 PT extension module for replacement of a reductive loop sequence, for
 PT producing polyketides, e.g. B1 avermectin -
 XX
 PS Disclosure; Fig 7a-f; 75pp; English.
 XX
 CC The present sequence is that of DNA encoding the first 2 modules
 CC of the avermectin polyketide synthase (PKS) of Streptomyces
 CC avermitilis. The invention relates to nucleic acids encoding a
 CC Type I PKS such as avermectin in which a polylinker with multiple
 CC restriction sites replaces or 1 more PKS genes encoding enzymes
 CC associated with reduction. Novel PKS are provided in which in
 CC which the reductive loop in a selected module of the Type I PKS is
 CC replaced with the equivalent segment from the same or different
 CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
 CC host cells, and methods for producing novel polyketides by
 CC culturing host cells are claimed. The polyketides obtained are
 CC useful as antibiotics and insecticides. Fermentation products
 CC containing C22-C23 dihydroavermectin, ivermectin and B1
 CC avermectins are claimed.
 XX
 SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;
 Query Match 100.0%; Score 534; DB 21; Length 12381;
 Best Local Similarity 100.0%; Pred. No. 4.3e-90;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGTCGGTGTGGTACCGGTGGACGGGTGCTGGGTGGCGGTGGCGGCATCTG 60
 DB 11426 GGGTCGGTGTGGTACCGGTGGACGGGTGCTGGGTGGCGGTGGCGGCATCTG 11485
 QY 61 GCTGGTGTGTGGGTGGGATCTGCTGTGTGAGCCGCGGTGGTCCGATGCTCG 120
 DB 11486 GCTGGTGTGTGGGTGGGATCTGCTGTGTGAGCCGCGGTGGTCCGATGCTCG 11545
 QY 121 GGTGGAGGGTCTGCGGGCGAGCTGCGCTGGGGGGGAGGTGCGGATTTGCG 180
 DB 11546 GGTGGAGGGTCTGCGGGCGAGCTGCGCTGGGGGGGAGGTGCGGATTTGCG 11605
 QY 181 TGTGATGTGGGGAGCGCGGAGGTGCTCGGCTGTGGAGGGTGTCTCCCGGGTGT 240
 DB 11606 TGTGATGTGGGGAGCGCGGAGGTGCTCGGCTGTGGAGGGTGTCTCCCGGGTGT 11665
 QY 241 CGGCTGACGGGTGTCTGTCATCGCGGTGTGTGTCGAGCATGCGACATGCTCTCTC 300
 DB 11666 CGGCTGACGGGTGTCTGTCATCGCGGTGTGTGTCGAGCATGCGACATGCTCTCTC 11725
 QY 301 ACGCCGAGCGGTGGGCAAGGTGTTCCGGCGCAAGGTGGATGCCGCTCTTTTGTGGAT 360
 DB 11726 ACGCCGAGCGGTGGGCAAGGTGTTCCGGCGCAAGGTGGATGCCGCTCTTTTGTGGAT 11785
 QY 361 GAGCTGACCGGGTATGAGCTGTGGGTTCGTGCTGTTCTCTCCGCGCGGGATC 420

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 106.227 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggctggtgtggtgacggg.....cgctggcgtggggcgtgtgg 534

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:.*
1: /SID82/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	534	100.0	11916	22	AAH79279
2	534	100.0	12381	21	AAZ58381
3	534	100.0	30690	21	AAH92301
4	534	100.0	30690	22	AAH79277
5	532.4	99.7	31422	21	AAH92302
6	532.4	99.7	31422	22	AAH79278
7	257.2	48.2	2700	22	AAD17193
8	257.2	48.2	65140	22	AAD17184
9	257.2	48.2	125401	22	AAD17186
					Streptomyces averm
					Streptomyces averm
					S. avermitilis ave
					Streptomyces averm
					S. avermitilis ave
					Streptomyces averm
					ERD48 insert DNA i
					Streptomyces nouns
					Streptomyces nouns

10	250.8	47.0	43280	18	AAT80413
11	250.8	47.0	4411529	22	AAI99682
12	249.2	46.7	4403765	22	AAI99683
13	244	45.7	27541	22	AAD17185
14	239.6	44.9	16767	22	AAF88339
15	239.6	44.9	29736	22	AAF88317
16	239.6	44.9	50000	22	AAF88313
17	239.6	44.9	80161	20	AAZ21501
18	230	43.1	14775	22	AAF88338
19	230	43.1	44377	18	AAT78508
20	230	43.1	44377	18	AAT80414
21	230	43.1	50000	22	AAF88316
22	223.2	41.8	20394	22	AAF24892
23	220.2	41.2	33529	23	AAI17367
24	218.8	41.0	4403765	22	AAI99683
25	218.8	41.0	4411529	22	AAI99682
26	212.4	39.8	15872	21	AAZ87283
27	209.2	39.2	15872	18	AAT68715
28	207.4	38.8	4851	22	AAH52062
29	200.6	37.6	53789	19	AAV21187
30	199.6	37.4	13842	21	AAZ87297
31	199.6	37.4	36778	21	AAZ87318
32	199.6	37.4	37948	21	AAZ87285
33	199.6	37.4	38506	21	AAA75633
34	199.6	37.4	38506	21	AAZ56001
35	199.6	37.4	47981	22	AAF30757
36	198	37.1	13987	18	AAT80415
37	197.6	37.0	5676	19	AAV21186
38	196.4	36.8	11220	21	AAZ87298
39	194.8	36.5	9513	22	AAF88337
40	192	36.0	29879	14	AAQ46806
41	183.6	34.4	68750	21	AAZ55887
42	182.4	34.2	6459	22	AAF88336
43	182	34.1	71989	21	AAZ29349
44	178.8	33.5	28598	17	AAT08769
45	178.8	33.5	28958	18	AAT89956

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX AC AAH79279;
XX AC
XX AC
DT 04-DEC-2001 (first entry)
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX Streptomyces avermitilis coding sequence derivative; AAS; avermectin derivative;
KW Avermectin aglycone synthase; AAS; avermectin derivative;
XX drug production; veterinary drug; pesticide; ds.
OS Synthetic.
XX Key Location/Qualifiers
FH 1..11916
FT /tag= a
FT /product= "AAG65268"
FT /partial
XX WO200162939-A1.
XX PD 30-AUG-2001.
XX PF 23-FEB-2001; 2001WO-JP01381.
XX PR 24-FEB-2000; 2000JP-0047405.
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

Tylactone synthase
Mycobacterium tube
Mycobacterium tube
Streptomyces nouns
S. spinosa DNA fra
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
S. spinosa DNA fra
platenolide syntha
platenolide syntha
S. spinosa DNA fra
Pimaricin biosynth
DNA sequence of S.
Mycobacterium tube
Mycobacterium tube
S. venezuelae vep
Streptomyces venez
Mycobacterium tube
Amycolatopsis medi
S. venezuelae macr
S. venezuelae pik
S. venezuelae pik
Nucleotide sequenc
Recombinant cosmid
Micromonospora meg
Hybrid strMG/tylg O
Amycolatopsis medi
S. venezuelae macr
S. spinosa DNA fra
Sorangium cellulos
Sorangium cellulos
Sorangium cellulos

THIS PAGE BLANK (USPTO)

Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001). Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/gbgpages/bgn/31/cover.html>)

crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Clouse laboratory at the University of California, Riverside (Choi, Clouse, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Segum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders/Also> see Clouse TJ, Wing R, Kleinbolls A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/wgn/31/cover.html>) "

stage subtraction cDNA library (hasp)"
/dev stage="Seedling roots of scots pine were infected for
6 days with H. annosum"
/note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
hybridization cDNA library was constructed from scots
pine roots infected for 6-days with mycelia of
Heterobasidion annosum (FPS)."
BASE COUNT 109 a 23 c 1007 g 4 t 79 others
ORIGIN

Query Match 1.0%; Score 299.6; DB 13; Length 1222;
Best Local Similarity 51.0%; Pred. No. 2.5e-32;
Matches 622; Conservative 0; Mismatches 593; Indels 4; Gaps 1;
QY 4512 CCAAAACCCACCAACCTGGCTCATCGCATCCCGGAAACCCAGACACACACCCACAT 4571
DB 1222 CCTCTCCCGCCCGCCCGCCCGCTCCCTCCCGCCCTCCCGCCCGCCCGCCCGCTCC 1163
QY 4572 CACCAACATCTTCAACCACTCCACCAACGAGCATACCCCGATCCCGCTTCAACCTCAA 4631
DB 1162 CCNNCCCGCCCGCCCGCCCGCTCCCGCCCGCCCGCCCGCTCCCGCCCGCTCCCGCTC 1103
QY 4632 CCACACCCACCAACCCCGACACCTCCACCAACGCTCCACCAACGAGCAACCAAGC 4691
DB 1102 CCCCCCCCCCGCCCGCCCGCCCGCTCCCGCCCGCCCGCCCGCCCGCCCGCTCCCGCTC 1043
QY 4692 CCAAAACACCAACCCGAGCATACACCGCGCTGTCTCCCTCTCGCCCTCGAGAAAC 4751
DB 1042 CCCCCCTTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 983
QY 4752 ACCCAACCCCGACCCCGCCACACCAACCGGACCTCTCAACCTCAACCTCAACCA 4811
DB 982 TCCCCCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCTCC 923
QY 4812 AACCAACCCCAACCCCGACCAACCCCGACCAACCCCGCTCTGTAGGACACCAACGAC 4871
DB 922 CCCCCCCCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 863
QY 4872 CACCAACCCCAACGACCCCTCAGACACCCCGACCAACCCCGGAGCTGGCGG 4931
DB 862 CNGCGCCCGCCCGCCCGCCCGCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 803
QY 4932 CACCAACCTCTCGAACACCCCGACCAACCGCGGAAATCATCGACCTCCCGACCAACCC 4991
DB 802 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 743
QY 4992 CACCCCGACACCTCCAGACCTCAGCAACCAACCTCAGCAACCGGACCAACCAACCA 5051
DB 742 CCGTCGCTCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCTCTCTCT 683
QY 5052 ACTCGCATTCGACACCGGACACCAACCGCGCGCTCAGCCGACCAACCTCAGCC 5111
DB 682 CCGCCCTCTCCCGCCCGCCCGCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 623
QY 5112 CACACCAACCAACCCCGACCCCGCCCGCCCGACCAACCTCATCAGCGGGAACCGG 5171
DB 622 CCGCCCGCCCGCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 567
QY 5172 CGCGCTCGGACCGACCTCAGCAACCTCAGCAACCGGACCAACCGGACCAACCTCTCT 5231
DB 566 CCGTCGCTCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCTCT 507
QY 5232 CCGTCACCAACCGGACCGGACCCCGACACCCCGCCCGACCAACCTCAGCCACCAACTCA 5291
DB 506 CCGCCCGCCCGCCCGCTTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 447
QY 5292 ACAAAAGGATCCACCTCAGCAACCTCGGACCAACCGGACCAACCGGACCAACTCA 5351
DB 446 TCTCCCGCCCGCCCGCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 387
QY 5352 ACAACTCTCAACAGATCCCGCCGACAGACCCCTCAGCAACCGGACCGTCAACCGGAGG 5411
DB 386 CCCCCCCCCCGCTCCCGCCCGCCCGCTCTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCG 327

QY 5412 CATCTCTCGACGAGCGGACCCCTCACCACCTCTACCCACCCCAACTCAACAAGCTCTCCG 5471
DB 326 CCGCT 267
QY 5472 CGCAGAAAGCCACAGGCGCCACCTCTCTCAACCAATCTCAACCAACACACCCCGCTCA 5531
DB 266 CCGCCCGCCCGCCCGCCCGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 207
QY 5532 CTTCT 5591
DB 206 TCCNCCCGCCCGCCCGCCCGCTTNGCGGTTCGTCGCCCGGCGGCGCCCGCCCGCC 147
QY 5592 CCGAGCAACAGCGCTACTCTGAGCGCTCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 5651
DB 146 CCGCT 87
QY 5652 CACCAAGCATCGCTGGGCGACCTGGCAAGAAACGAGCTCTGTGATTCGGAAGCGCCG 5711
DB 86 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 27
QY 5712 CGCATATCTCGACCGCGCG 5730
DB 26 CCGCCCGCCCGCCCGCCCGCCCG 8

RESULT 10
BF256755/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF256755 1196 bp mRNA linear EST 22-OCT-2001
HVSMEF0010017f Hordeum vulgare seedling root EST library HVCDA0007
(Etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0010017f, mRNA sequence.
BF256755
BF256755.2 GI:16315528
EST.
Hordeum vulgare.
Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1. (bases 1 to 1196)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
, R.D., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex unstressed seedling root cDNA library
Unpublished (2001)
On Nov 16, 2000 this sequence version replaced gi:11185868.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7289
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 182
Seq primer: AATTAACCTCTACTAAAGGG
High quality sequence start: 95
High quality sequence stop: 1129.
Location/Qualifiers
1. 1196
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0010017f"
HVCDA0007 (Etiolated and unstressed)"
/tissue type="Seedling root"
/lab host="TJ121"
/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered

FEATURES
source

RESULT 7
BH146899/c
LOCUS
DEFINITION
ENTOM477F Entamoeba histolytica Sheared DNA linear GSS 27-AUG-2001
genomic, DNA sequence.
ACCESSION
BH146899
VERSION
BH146899.1 GI:15302995
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
1 (bases 1 to 1222)
AUTHORS
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL
Unpublished (2001)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 452
High quality sequence stop: 563.
Location/Qualifiers
1. .1222
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 125 a 0 c 1048 g 49 t
ORIGIN

Query Match 1.0%; Score 306.4; DB 17; Length 1222;
Best Local Similarity 55.4%; Pred. No. 2.8e-33;
Matches 654; Conservative 0; Mismatches 521; Indels 5; Gaps 3;

QY 4489 CCCTCACCCTCCACACCCACCAACCAACCAACCTGGCTCATGCGCATCCCGGAA 4548
DB 1179 CC 1120
QY 4549 ACCAGACCCACACCCCCCAGATACCAACATCTTCCACCAACCTCCACCAACGCGCATC 4608
DB 1119 CCGCCTCTCCACACCCCGCCCTCCCTCTTCCCGCCCGCCCTCTCCCGCCCGCCCA 1060
QY 4609 ACCCCATCCCTCCTACACCAACCAACCAACCAACCAACCTCCACCAACCC 4668
DB 1059 CTCCTCCCTCCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1000
QY 4669 CT--CCACACACCCGACAAAGCCCAAAACACACACCGAGGAGCATCACCGGCTGC 4726
DB 999 CTTCCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 940
QY 4727 TCTCCTCTCTGCGCTCGAGAAACACCCACCCACCCACCCACCCACCCACCCACCGGA 4786

DB 939 CC 880
QY 4787 CCCTCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCT 4846
DB 879 CC 820
QY 4847 GGTACGCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4906
DB 819 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 760
QY 4907 AAGCCCAAACTGGGAGCTGCGCGGACCAACCTCTCTGAAACACCCACCAACCGCG 4966
DB 759 CTCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 700
QY 4967 GAATCATGACCTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5026
DB 699 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 640
QY 5027 TCACCAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCGCC 5086
DB 639 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 581
QY 5087 GCTTACCCCGGACCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCT 5146
DB 580 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 521
QY 5147 CCACCTCATCAACCGGCGGAAACCGGCGCTCGCCACACCAACCAACCAACCAACCAACCA 5206
DB 520 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 461
QY 5207 CCAACCAACCCCAACCAACCTCTCTTCCACGAGCGGAAACCGGCGCCCAACCAACCAACCA 5266
DB 460 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 401
QY 5267 CACAACCTCAACCAACCAACCTCAACCAACCAACCAACCAACCAACCAACCAACCTGCG 5326
DB 400 TCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 341
QY 5327 ACACGAGCAACCAACCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCAACCAAC 5386
DB 340 TTTTCTCTTCT 281
QY 5387 TCACCAACCTCAACCAACCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAAC 5446
DB 280 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 221
QY 5447 CCACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCAAC 5506
DB 220 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 161
QY 5507 TCACCAACCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAAC 5564
DB 160 CCGCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 101
QY 5565 CCGCGCACCCCGGCAAGCAACCTAGCGCGGACCAACCTCAACCAACCTCAACCAACCTCAACCA 5624
DB 100 TATTTTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 41
QY 5625 CCACCGGCAACCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCAACCAACCTCA 5664
DB 40 ACCACACCCCGGCAAGCAACCTAGCGCGGACCAACCTCAACCAACCTCAACCAACCTCAAC 1

RESULT 8
AW949571/c
LOCUS
DEFINITION
EST361641 MAGE resequences, MAGA Homo sapiens cDNA, mRNA sequence.
ACCESSION
AW949571
VERSION
AW949571.1 GI:8139202
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
Db 712 CACCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 653
Qy 4849 TAGGCAACCAACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 4908
Db 652 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 593
Qy 4909 GCCCAAACTGGGAGCTCCCGCGCACACCTCTCTGAAACACCCACCAACCAACCAACCAACCA 4968
Db 592 CCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTAC 533
Qy 4969 ATCATGACCTCCCAACCAACCAACCAACCAACCAACCAACCTTCCAGACCTTCAACCAACCTTC 5028
Db 532 CCCCCCCCCCAACCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 473
Qy 5029 ACCCAACCCCAACCAACCAACCAACCAACCTGCGCATCCGACCAACCGGACCAACCAACCGCGC 5088
Db 472 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 413
Qy 5089 CTCACCCCCCAACCAACCTTCAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5148
Db 412 CCCCCCCCCCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCT 353
Qy 5149 ACCTCATCAGCGGCGGAACCGCGGCTTGGCACCTTCAACCAACCAACCTTCAACCAACCTTCA 5208
Db 352 NCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 293
Qy 5209 CACCAACCAACCAACCAACCTTCTTCAACGAGCGGACCGGCCCCCAACCAACCAACCAACCAACCA 5268
Db 292 CCCCCCNCCCCCTTCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 233
Qy 5269 CAACACCTCACCAACCAACCTTCAACCAAAAGGATCCACCTTCAACCAACCAACCAACCTTCCGAC 5328
Db 232 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 173
Qy 5329 ACAGCAACCCAGACCAACCTTCAACCAACCTTCTTCAACCAACCAACCTTCAACCAACCAACCTTC 5388
Db 172 CCCCCCTGCCCCCCCCCCCCCAACGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 113
Qy 5389 ACCACCGTATCACACCGCAGGATCTTGGAGGACGACCACTTCAACCAACCTTCAACCAACCTTCA 5448
Db 112 CCCCCCCCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 53
Qy 5449 ACCCAACTCAACCAACCTTCTTGGCGGCAAAAGCCACAGCGGCCCAACCTTCTTCC 5500
Db 52 CCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1
```

RESULT 4

```
BI416573/c
LOCUS
DEFINITION
hasp001xgl1f Heterobasidion annosum - Scots pine infection stage
subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp001xgl1f, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

BI416573 1223 bp mRNA linear EST 15-AUG-2001
Pinus sylvestris/Heterobasidion annosum.
Eukaryota; mixed EST libraries.
1 (bases 1 to 1223)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:08:39 ; Search time 24141 Seconds
(without alignments)
20589.018 Million cell updates/sec

Title: US-09-914-286-1
Perfect score: 30690
Sequence: 1 gtgcagagatgacgcggg.....acgaagtgagacagcgtga 30690

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
EST:*			
1:	em_estba:*		
2:	em_esthum:*		
3:	em_estin:*		
4:	em_estmu:*		
5:	em_estov:*		
6:	em_estpl:*		
7:	em_estro:*		
8:	em_estc:*		
9:	gb_est1:*		
10:	gb_est2:*		
11:	gb_est3:*		
12:	gb_est4:*		
13:	gb_est5:*		
14:	gb_est6:*		
15:	em_estfun:*		
16:	em_estom:*		
17:	gb_gss:*		
18:	em_gss_hum:*		
19:	em_gss_inv:*		
20:	em_gss_pln:*		
21:	em_gss_vrt:*		
22:	em_gss_fun:*		
23:	em_gss_mam:*		
24:	em_gss_mus:*		
25:	em_gss_other:*		
26:	em_gss_pro:*		
27:	em_gss_rod:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	345.6	1.1	1197	13	B1416470
C 2	335.2	1.1	1224	13	B1416537
C 3	329	1.1	1015	13	B1416550
C 4	327.8	1.1	1223	13	B1416573
C 5	323	1.1	1186	12	B1416575
C 6	310.4	1.0	931	13	B1416575

C 7	306.4	1.0	1222	17	BH146899
C 8	301	1.0	1453	10	AW949571
C 9	299.6	1.0	1222	13	B1416660
C 10	299.2	1.0	1196	12	BF256755
C 11	298.2	1.0	1042	13	B1416535
C 12	297.6	1.0	1082	10	BE455154
C 13	297.6	1.0	1251	10	BE455089
C 14	295.6	1.0	1166	12	BF256751
C 15	295.2	1.0	1005	17	BH148610
C 16	293.4	1.0	1138	17	AG152606
C 17	292	1.0	1321	17	AG126084
C 18	291	0.9	1138	12	BF256506
C 19	290	0.9	1189	10	BE455091
C 20	289.2	0.9	873	12	BG786249
C 21	286.8	0.9	1277	12	BG447302
C 22	285.6	0.9	1606	13	BM470613
C 23	283.8	0.9	1161	14	BM912505
C 24	282.2	0.9	1076	17	AG032994
C 25	282.2	0.9	1161	12	BF256588
C 26	282.2	0.9	1195	13	B1416506
C 27	282.4	0.9	1303	14	BM908878
C 28	281.8	0.9	1152	12	BF256631
C 29	280	0.9	1328	10	BE455227
C 30	279.2	0.9	1235	13	BM475329
C 31	278.8	0.9	1177	17	AG081646
C 32	278.6	0.9	1294	12	BF256749
C 33	277.6	0.9	1116	12	BF256617
C 34	277.2	0.9	1184	14	BM911719
C 35	277.4	0.9	1417	12	BF257444
C 36	276.8	0.9	1297	14	BQ064843
C 37	275.8	0.9	1088	12	BF256580
C 38	275.2	0.9	1137	17	AG078502
C 39	274.6	0.9	1293	12	BF256498
C 40	273.4	0.9	877	13	B1416653
C 41	273.6	0.9	1292	12	BF255049
C 42	273.2	0.9	1391	10	BE455155
C 43	272.8	0.9	1036	12	BF254665
C 44	271.6	0.9	1201	14	BQ063878
C 45	270.2	0.9	960	14	BQ429353

ALIGNMENTS

RESULT 1	BI416470	BI416470	1197 bp	mRNA	linear	EST 15-AUG-2001
LOCUS	BI416470/c	hasp001xa01f	Heterobasidion annosum - Scots pine infection stage			
DEFINITION		subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.				
ACCESSION	BI416470					
VERSION	BI416470.1	GI:15187493				
KEYWORDS	EST.					
SOURCE	Pinus sylvestris/Heterobasidion annosum					
ORGANISM	Pinus sylvestris/Heterobasidion annosum					
REFERENCE	1 (bases 1 to 1197)					
AUTHORS	Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.					
TITLE	Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)					
JOURNAL	Unpublished (2001)					
COMMENT	Contact: Fred O. Asiegbu Dept. of Forest Mycology & Pathology Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden Tel: +46 18 67 15 98 Fax: +46 18 30 92 45 Email: Fred.Asiegbu@mykopat.slu.se Seq primer: T7 primer. Location/Qualifiers 1. .1197 /organism="Pinus sylvestris/Heterobasidion annosum"					
FEATURES	source					

Db	6462	 CGTCGTGGCCACTCGCAGGCGAGATCGCGCGCGGTACGTGCGCGGTGCCCTGAGCCT	6521
Qy	26733	GAAGAAGCGCGCAAAACCGTTGCACATGCGCAGCGCGCACATGGCGCGC---CGTACGAGG	26789
Db	6522	GGACGAGCGCGCTCGTGTCTGTGTGACCTTGCACAGAAAGTCAATCGCGGCGCCACTCGCGGG	6581
Qy	26790	CCGGGGGCCCATGGCCCTCACTGCGCCCTGCGCGGCCACGAGACGTGCAGCAGCTCATTTCCGA	26849
Db	6582	CCAGGCGGCATGCTGTCCCTCGCGCTGAGCAGGCGCGCGTGTGTGAGCGCATGGCCG-6640	
Qy	26850	ACGGTGGGAAGGGAGTGTGGGTGGAGCCCTCAAAGCGGCCCACTCCACCAACCGTCTC	26909
Db	6641	-----GGTTTCGACGGGCTGTCCGTGCGCGCGGTCAAACGGGCGTTACGCCCAACGCTGTTTC	6695
Qy	26910	CGGCGACACCAAGCGGTGTGATGAGTGTGGCGCACTGCACCGACACCGGCTCAGGGC	26969
Db	6696	GGGCGACCCGACCCAGATCCAAAGCTTCGTTCAGGCGTGTGAGCGCCGACGGGGTCCGGC	6755
Qy	26970	CAAAACGATCCCGTGCAGTACGCTCCCACTGCCCGCCACGTCCAAACCCCTCCACGAGGA	27029
Db	6756	ACGGATCATCCCGTGCAGTACGCTCCCAAGCGCCACAGCGCCACGTGAGACCATCGAGAGCA	6815
Qy	27030	ACTCTGCACCTGTGGGAGACATCAACCCCGACGCGTCCACCGTGCAGTGCCTTCTTCAC	27089
Db	6816	ACTGCCGACGTCTTGGCGGGTGTGCCCGCCAGACACCCAGGTCCCTTCTTCTCCAC	6875
Qy	27090	CGTGGAGGCACCTGGCTGGACACCAACCTTGGACGCGGCTACTGTACCGCAACT	27149
Db	6876	CCTGGAAGGCGCTTGGATACCGAAACCGCGCTTCGACGCGGCTACTGTGTACCGCAACT	6935
Qy	27150	CCACGACCGCTCCGTTTCAGCCACGCCATCCAGACCCCTG---ACCGACGACGACACCG	27206
Db	6936	CGGCATCGTGTGGCTTCGCCCCGCGCGTGGAAACCTTGGCCACGAGCGAGGCTTCAC	6995
Qy	27207	CGCCTTCATCGAAATCAGCCCCCAACCCACCTCTGTCGCCGCAATCGAAGACACCAACCGA	27266
Db	6996	CCAATTCTGTCGAGGTGAGCGCCCAACCCGCTCTCTCAACATGGCGCT-----7040	
Qy	27267	AAACACACCGAANAATCACCGCGACGGCAGCCTCTCGCGGGGGACACGACACCA	27326
Db	7041	-----GCCCGAGACCGTCAACCGGATCTGGGCACTCGCGCGGTGACAAACCGCGGACGCA	7094
Qy	27327	CGGCTTCTCACCGCCTCGCCCAACCCACACCAACCGGCACTCGGACACCCACCAACCTG	27386
Db	7095	CGGCTCAACACTCTCTCGCGAGGCTGGGCGCAAGG-----7133	
Qy	27387	GCACGACCACTACACCCAAACCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAAC	27446
Db	7134	---CCTCACCGTGCAGTGGGCGCTCTCTCTCCGCCACACGACCAACCCCGATCTGCC	7190
Qy	27447	CACCTACCCCTTCCAACACAGCACTACTGGTCCCAAC---ACCCACCAACAACACCA	27503
Db	7191	CACCTACGCTTCCAGACGAGCGCTACTTGGCGGAGCGCGGCGGCGGCGGCGGCGGCGG	7250
Qy	27504	CCTCACACCAACCGGCTCACCCCAACCAACCCCTCTCTCAACCGCACTCACCTC	27563
Db	7251	CATCACTCGCGGCTCTCGGGGCGCGAGCAACCGGTGTCTCGGGGCGGCGGCGGCGGCT	7310
Qy	27564	CGCGCAACAACAACAATACTACTACCGGCGCGCTCTCCCTACGCAACCCACCCCTGGCT	27623
Db	7311	CGCGGCACTCGACGGGTGTGTCTACAGGGGAGCGCTCTCCCTCCGTACGCAACCCCTGGCT	7370
Qy	27624	CACGACCAACCGTTCGCGGAGTGTCTCTCTGCGGGGCAACCGGCTCTCTGAACTCGC	27683
Db	7371	GGCGGACCAACGCGGTGGCGGCAACCGTGTCTGTGCGGGAAACGGGCTTCTGGGAGCTGGC	7430
Qy	27684	CCTCCAAGCGCGGAAACGGGTGACTCGCCTCTCGGTTGGAGGAATGAACTCTGCAACCGC	27743
Db	7431	GTTCGAGCGGGACCAAGTGGTGTGATCTGTGTGAGGAGCTCACTCTCGAGCGGCC	7490
Qy	27744	GTTGGTGTATCCCGCAACACGAGGACGTGTGAGTGTGAGGTCAACGTTTCGGGCGAGCCGATGA	27803

QY	24513	CGTTCTCGGCTTTCGTAATCCGAGGCGATCGAGGACCAACGAGCCTTCCGCGACTCGG	24572
Db	4371	GGTCTCTCAACACCCCTCCCGAGGCGCTGCACACGCGGGCGGGCTTCCGTGACCTCG	4430
QY	24573	CTTTCGACTCGCTGACCTCGGCTCAGTTTCAAGCAAGAACTCGCAAGGAAACCGGACTGCC	24632
Db	4431	ATTTCGACTCGCTGACGCGGCTCGAGCTCCGCAACCGCTCAAGAACGCCACCGGCTGGC	4490
QY	24633	ACTCCCCCGTCTCGCTTTCGACTATCCACCCCGCGAGGAATGTGCTGCCCATCTGGG	24692
Db	4491	CTTCCCGGCACTCTGCTCTTTCGACTACCCGACCCCGCGAGCGTGGCGGAGTTCCTCT	4550
QY	24693	CACACAACTCGTGCACCTAGACGACGAAGAGAGCGGCACTGTGGAATGCTCTCCGCA	24752
Db	4551	CGGGAGATCTTGGGCGAGGAGCGGCTGCGGCGAGCAGCTTCGCTGAGCGGGG--	4608
QY	24753	AGTGGCCATCGGCGTACCGTTCGAGGACGAACCGATCGGCATCATCGTATGCGCATGTG	24812
Db	4609	-----GTGACGACGAGCCCGTCCGATCGTGGCATGGCGTGGCG	4649
QY	24813	CTTCCCGGCGGCTAGCTTTCGCCGACGACCTGTGGGAATTTGCTCGCTTCGGGTAA	24872
Db	4650	CTTCCCGGCGGCTGCTCGCTTCGCCGAGGACCTGTGGCGGCTGGTGGCCGGCGGAGGA	4709
QY	24873	CGCTATCGGCGTCTTCCGACCGACCGCGCTGGGACCTGGACAGCGTCTTACGACCCGA	24932
Db	4710	CGGATCTCTCGGCTTCCCGAGGACCGCGCTGGGACGTGGAGGGCTGTACGACCCGGA	4769
QY	24933	CCCCGACCAACCCCGGCACTGTCTACACCCGAAACGGCGGATTCCTCTACGGGCGAGGCA	24992
Db	4770	CCCGGAGCGTCCGGGCGAGCTACTGCGCTGCGGCTTCTTCGACGAGGCGGGCA	4829
QY	24993	CTTTCGAGCGCGAATTTCTTGGGATCAGCCCGCGGAAAGCCCTCGCCATGGAACCCGAGCA	25052
Db	4830	GTTTCGAGCGCGAATTTCTTGGGATCTCGCGCGGAGGCGCTCGCCATGGAACCCGAGCA	4889
QY	25053	ACGACTCTCTCGAAACCGCTTGGGAACCATCGAAACCGCGGATCAACCCCGACAC	25112
Db	4890	CGGCTCTCTCTGGAGACCTCTCTGGAGGCGCTCGAGGACCGCGGATGACCCGACCTC	4949
QY	25113	CTTTCAGCGGACCCCGGAGTCTTTCGCGGAAATCAAGCTCAAGACACCGCCGCA	25172
Db	4950	CTTTCAGGGGCGAGCTGCGGCTGTTTCGGGSCACCAACGGGCCCTTACGAGCGCT	5009
QY	25173	TATCCGCCAAGCGGTGATGAGACCATCGAGGCTAGCCCTCAGCGGAGTTTGGG	25232
Db	5010	GCTCCGCAACACC-----GCCAGGATCTTGAAGGTTTACGTCGGGACGGGCAACGCCG	5063
QY	25233	AGTGTGGCTCGGCGGCTGACGCTCGGCTCGAGGCGCCCGCGTGTGCT	25292
Db	5064	CAGCATCATGTGGGCGGTCTCTGTACACCTTCGGCTGGAGGGCCCGGCTCACGCT	5123
QY	25293	GGATACGGCGTTCGTGCTGTGTGGGTGCAATGGGCGCGAGCGTTCGCTGC	25352
Db	5124	CGACACCGCTGCTCTCTCGTGTGCGCTCGCCCTGCACCTCGCGGTGACGCGCCCTGGCAA	5183
QY	25353	GGGTAGTGTTCATGCGCTTCGCGGGGTGACGGTGTGCTCTCGGGTACGTT	25412
Db	5184	GGGCGAATGCGGCTGCGGCTCGCGGCGGTGTGACGGTCAATGACGCCCAACGACGTT	5243
QY	25413	TGTGGAGTTCTCACGTACGGGGGTCTGGCCGCGGACGGCGGTGCAAGGCTTATTCGGC	25472
Db	5244	CGTGGAGTTTCAGCGGAGCGGGCTTCGCGAGGAGACGGCGGTGCAAGGCTTCGCGC	5303
QY	25473	GGCTGTGACGTTACCGGCTGGGCGAGGTTGGGGATGCTGCTGGTGGAGCGGCTCTC	25532
Db	5304	GTCCGCGGACGGCTTCGGCCCGCGGAGGCGTCCGATGCTCTCTGTGAGCGCTGTC	5363
QY	25533	CGACGCGGTCGCAACGGTCAACGCTGCTGGCGGCTGGCTGGGAGTGGCATGCGTCAACCA	25592
Db	5364	GGACGCGCGCGCAACGGACACCGTGTGTGGCGGCTGTGTCGCGGCGGCGGCTCAACCA	5423
QY	25593	GGACGCTCGAGCAACGGTCTGACCGCGCCCAACGGGCGCTTCCAGCAGCGCTGATCCG	25652
Db	5424	GGACGCGCGAGCAACGGCTTGAACCGCCCGAACGGGCGCTTCCAGCAGCGCTGATCCG	5483
QY	25653	TCAGGCGCTGGCCAAATCGGGACTGACCCCGCGATGTTCAGCAGTGTGAGGGCGACGG	25712
Db	5484	GCGCGCTCGCGGACCGCCGACTGACACCGCGCAACGTGAGCGCTGTGAGGGCGCACGG	5543
QY	25713	CACCGGACCACTCTGGGGGACCGGATCGAGGCGCCAGGCACTCTTGGCGCGCTACGGACA	25772
Db	5544	CACGGGACCGGACTTCGGGACCGGATCGAGGCGACAGGCGCTTATCGGCCACTACGGCCA	5603
QY	25773	ACACGCGCCCAACACCGCCCTTGTGGCTGGGATCCCTCAAAATCCAAATCGGGCACGC	25832
Db	5604	GGGGCGGACACACGACGCGCTGCGCTTGGGCTGTTGAAGTCCAAATCGGACACAC	5663
QY	25833	ACAGGCGCGCGGGGCTGGGCGGAGTCAATCAAGATGTGTATGGCCCTTGGGCAACGGCT	25892
Db	5664	CAAGGCGCGCGGCTGTCTCGGCTATCAAGATGTTCAGGCGATGCGGCCACGGCGT	5723
QY	25893	GCTGCCACAGACCTTCCACGTTGGAGCGCCACCCCGAGGTCGACTGCTTCCACAGGCGC	25952
Db	5724	CCTCGCAAGACGCTTCCACGTTGGACCGGCGCTGCGACCGATCGACTGCTGGCGGGCAC	5783
QY	25953	AGTCAAACTCTTGACACACCGGCTGCGCTGCGCCCGACCGCGCGCGCCACCGCA	26012
Db	5784	GGTGCAGTGTCTACCGAGGCTATGGAATGCGCGGAGGAGGAGGCGGGCTGCGCGC	5843
QY	26013	GCGCGCTGTGATCATATTGCGGCTGAGCGGACCAACCGCCACATCATCTCTGGAAGAAGC	26072
Db	5844	GCGCGCTGTCTCTCTTTCGCGATCAGCGGCAACGACCGCACATCTGTCTGGAAGAAGC	5903
QY	26073	ACCACTCCCGAGGACGATACCGACGAGAACCGCTTCCCAACGACGACGCGCTGCC	26132
Db	5904	CCC-----GGTGCAGAGAGACGCCCCCGGCGAGAGCGCTGCGGTGCG	5945
QY	26133	CCATCCCTCTCTTCCGCTGCGGCTGCGCGAGGCTGAGGCGCGGTTGCGGGCGCA	26192
Db	5946	CGGTGTGTGCGGTGCTGCTCGCGAAGACTCCGCGCGCGCTGGAGCGCCAGATCGG	6005
QY	26193	GGCAGAGGTTTGGCCAGTGTGTCGAGCGCGCGCGCGGACATGTCACTTGGGACATTTGG	26252
Db	6006	ACGCTCTCGCGGTTTCTGCTCGAGGCGCTGACGAGCGCGCGGATCCGCGGCGGCTGCG	6065
QY	26253	TGCGGCTTGGCGCGCGCGCGCTTCTGGAACACCGCGCGCTCATCTTGGCGCGCA	26312
Db	6066	TGCGCTACTTGGCGCGCGCGGCTGTCGAGTTTGAAGCACCGGCGGCTGCGGCTCGGACCGG	6125
QY	26313	CCGCGAGGAACTGGCGGAGGCACTGACAGCCCTGGGACCGCGGCAACCCCAACAT	26372
Db	6126	ACAGGACGACCTGGCGCGGCTGCGCGCGCTGAGG-----	6163
QY	26373	CACCAAGGCGCACACCGCGGCGGTGACCGGCGGCGGTGCTTCTTCTTCCCGGACA	26432
Db	6164	--GTCTGTGTCGCGGTGTGGCTTCCGCTGAGTGTGAGTGTGCTTCTTCTTCCCGGACA	6221
QY	26433	GGGCGCGAGTGGCGGCGGATGGGCGCTGACCTCTGCTCACCGCTTCCCGGCA	26492
Db	6222	GGGACGAGTGGCGGCGGATGGGTCGCGATCTCTGACGCTGCAAGGAGTTTCCGGC	6281
QY	26493	ACACATGACGATCGAGGAAAGCCCTCACCCCTTGGGTGCGCTTCTTCTGACCGACAT	26552
Db	6282	GGCCATGGCCGAGTGGAGGCGCGCTGCTTCTGCTGAGTGTGCTGCTGAGGCGCT	6341
QY	26553	CTTTCAGCGGACCCCGGACGACCCCGCATGGCAACGCGGAGTGTGTCAGCGCGTGT	26612
Db	6342	CGTCCGACAGGCGCGCGCGCCCAACGCTGGAGCGGCTGATGTCTGCTCAGCGCGTAC	6401
QY	26613	CTTTCAGCATGCTTCTCTCGCGCGCTTGGGCTTCTTACGCGATCGGCAACCGGACGC	26672
Db	6402	GTTTCGCGCTGCTGCTGCGGAGAGGCTTGGGAGCAACGCGGCTGACCGCGGAGC	6461
QY	26673	GCTCTCGGCCACTCCCGAGGAGAAATCGCGCGCGCCACATCTTGGCGGCGACTCAGCCT	26732

Db 2376 CTTTCTGACACTCGAAGGCGCTTGATFACCGAGCGGTGCTCGAGCGCACTACTGGTA 2435
QY 22404 CCGCAACACCGTGCAGTACGCCACCAACCCAAACCTCCACCAACAGCGGTGACCAAC 22463
Db 2436 CCGCAACCTCGCAACCGGTGGCTTCGCCCGCGGTGAGAACCTCGCCACCGACGA 2495
QY 22464 CTATATCGAACTCGGACCGGACAAACCTTCACCAACCTCACCACCAACCAACCTCCCAA 22523
Db 2496 AGGCTTCAACCACTTCATCGAGTTCAGCGCCACCGCTCTCAACATGACCTCCCGA 2555
QY 22524 CACCCCAACACACCTTCACCTTACCTTACCTACCCACCGCCACCAACCCCAACCACTCTCT 22583
Db 2556 GACCGTCAACCGGCTCGGCAACCT---CCGCCGAACAGGAGGCGAGGAGGTCTGTT 2612
QY 22584 CACCAACCTCGCAAAACCAACCAACCTTGGACCCCAACCTTACACCTTACACCCACCAACCAAC 22643
Db 2613 CACTCACTCGCGAAGCTTGGACCAACGCGCTTCAACATGACTGGGCGCGCTCC---T 2669
QY 22644 CCAACCCCAACCAACCAACCTTGCAGCTTCCCACTTCCCACTTCCCAACCAACCAACCACTA 22703
Db 2670 CCCCACGCAACCGGCAACCAACCCCGAGCTTCCCACTTCCAGGCGGTCACTA 2729
QY 22704 CTGGCTCGAACTACCCAGCGCCCAACCAACCGGCTTAAAGGCTTCTGCGGCTCGGC 22763
Db 2730 CTGGCTCCA----- 2738
QY 22764 TCCAGACACCGCGAGTTCGGAGTTCTGGAGCGGTGACGAGGAAGACTTCAGAGCT 22823
Db 2739 ----- 2738
QY 22824 CGCGAAACCTCGAATGACGCTCTGTCTGGACAGGTGGTGGTCCGCACTCTCCGC 22883
Db 2739 ---CGACTCCCGCGTCCAGGCTCCGTGCAGACT----- 2773
QY 22884 CTGGCACCGCAACCAACGACCAAGCGCGCATCAACACTTGACCTACAGGAACCTG 22943
Db 2774 -----CTGGCGCTACCGCATCGACTG 2795
QY 22944 GAAACCTTCACTCCCGACCAACCAACCAACCTTCACTCCGCTATCGCCAT 23000
Db 2796 GAAGCGCTTCGGGTTCGGACGCTCGGAGCGCGCGGTGTCGGGCGGTGGGTCT 2855
QY 23001 CCCCAGAACCCAGACCAACCAACCCCAACATCTTCAACAACTTCCACCAACCA 23060
Db 2856 CGTGTCTCCCGAGGACCGTTCCGCGGAGCGCCCGGTGCTCGCGCTGTCCGCGC 2915
QY 23061 CGGCATCAACCCCATCCCTCTACTGTCAACCAACCAACCAACCAACCAACCAACCTCA 23120
Db 2916 CGCGCGCAACCCGTAACGTGGACGTGTC-----CCCGCTGGGCGACCG 2960
QY 23121 CCACACCTTCCACCAACCGGACCAACGCAACCAACCAACCAACCGGACCTTACCGG 23180
Db 2961 GCAGCGCTTCGGCGGACCTTGGGAGGCGCTTGGCGGCGCGGTGGAGCGCTCGACGG 3020
QY 23181 CTGTCTCTCTCTCTCGCTTCGAGAAACACCCACCCCAACCAACCCCAACCAACCAAC 23240
Db 3021 CGTCTCTCTCTCTCTCGGTGGAGAGAGCGGCAACCCCGGCAACCCCGCCCTTAC 3080
QY 23241 C---GGCAACCTCTCAACCTTCACTTCCCAACCAACCAACCAACCAACCAACCAAC 23297
Db 3081 CGGGGCAACCGGCGCACCTTCACTTGTGAGGCGGTGGAGGAGCGCGGCTCGCGC 3140
QY 23298 CCCCCTCTGTAGGCAACCAACCGCAACCAACCAACCAACCAACCAACCAACCAACCAAC 23357
Db 3141 CCGCTGT 3200
QY 23358 CCCCACCAAGCCCAACCTGGGAGTTCGCGCGCAACCAACCTTCTTCAACCAACCAACCAAC 23417
Db 3201 CCCCAGGAGGATGT 3260
QY 23418 CACCGCGGAATCATGACCTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCT 23477

Db 3261 GTGGGCGCGCTTATCGACCTGCTTGGACGCGGACCGGCGCGCTTGGACCGCATGAC 3320
QY 23478 CCAAAACCTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 23537
Db 3321 CAGGCTCTTGGCGGCGGTACGGGTGAGGACCAAGTTCGGGTACGGCTTCCGGGTGCT 3380
QY 23538 CACCGCGCGCTTACCGCCCAACCAACCTTCAACCAACCAACCAACCAACCAACCAACCA 23596
Db 3381 CGCGCGCGCTTCTTGGCGGCTTCCCTCCCGGCGACGCGCTTCCCGGTGGGCA 3440
QY 23597 ---CCACGGAACCAACCTTATCAACCGGCGGAAACCGGCGGCTTCCCAACCAACCA 23654
Db 3441 GCGCGAGCGACGCTGCTGCTCAACCGGTCCGAGGAGCTGCGGCGCGGAGCGCGACG 3500
QY 23655 CCACCTTCAACCAACCAACCAACCAACCTTCTTCAACCAACCAACCAACCAACCAACCA 23714
Db 3501 CCGGCTTGGCGCGGAGCGGCGGAGACCTCTCTCTCAACCAACCAACCAACCAACCAACCA 3560
QY 23715 CACCGCCCAACGCAACAC---CTCACCACCA 23744
Db 3561 AGCGCGCGAAGGACCTTCCGCTGCGCGGAGGACTCCGCGCTTCCGCGGCTCGTCCCGA 3620
QY 23745 ACTCCAAACAAAGGATCCACTCAACATCAACCTTGGAGACCAAGCAACCAAGACCA 23804
Db 3621 ACTCGCGGACCTTGGGCGGAGCGGCAACGCTGCTGACCTGCGACCTCAACGAGCGGAG 3680
QY 23805 ACTCCAAACAACTTCAACCAACCAACCAACCAACCTTCAACCAACCAACCAACCAACCA 23864
Db 3681 GCGCGCGCGGCTGCTGCGGCGCTTCCGAGCGGACCGGCTTCCGCGGCTCTCCACCT 3740
QY 23865 CGCAGCGCTCAATCTTTCGCGCGGCTTGGAAACCGATGCGGAATCTTCTTTCCT 23924
Db 3741 GCGCGCCACCGTGCAGTCCGAGCGCTTCCGCGGAGCGGAGCGGCTCGCGGCTGT 3800
QY 23925 TACGCGAGGAGGCAACGCGGCGGCGATTTCTGATGATGTTGCTGTGACCATGAAC 23984
Db 3801 CGTGACCGGAGGCGGAGCGGCGGCTCCACCTGAGCGGCTTCTTGGGAGGCGCGCGC 3860
QY 23985 GCTTGA-----CACTTCACTTCTTCTGCGGCGGCGGCGGCTTGGGCGAG 24032
Db 3861 TCGCGAGCGGCTTCCGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3920
QY 24033 CGGGAATCAGTTCGCTACTTCCGCGGCAACGACATACCTTGAACGCGCTTCCGAGCATCG 24092
Db 3921 CGCGGCTCAGGCGGTACGCGCGGCTTACGCGCTTCTTCTGAGCGGCTTCCGCGGTGAG 3980
QY 24093 TCAGACATGACATTCCCGGCGGATCGATCGCTTGGGCGGCTTGGGCGGAAAGGAT 24152
Db 3981 CGGGCGGACGCGGCGGACCGGTGACCTCGGTGGCTTGGAGCGGCTTGGAGGCGGCGGCT 4040
QY 24153 GTGCGCGGCTGATGCGGCTCATGTTTACCTGGAAGCGCGGCTTCTGCGATGGAGCC 24212
Db 4041 CACCGAGGCTGCGACCGGAGCGG---CTGCGCGCTTCCGCGCTTCCGCGCTTCCGCGC 4097
QY 24213 ACGATGCGGCTTCCGCGGATTCATTCGTCGCGGCGGAGCGGCGGCTTCCAACTGAT 24272
Db 4098 CGCGAGCGGCTTACCGCGCTTGGACACCGCGCTTGGCGGAGGAGGAGGAGGAGGAGGAG 4154
QY 24273 CATCGGACATTCGACTGGGAGGCTTCTTCCCGGCTTCAACGCTTCAACGCGGAGGAGGAG 24332
Db 4155 GATCGCGGAGCTGAGTGGTGGCTTCCCGCGGCTTCCACCGGCGGCGGCGGAGGAGGAG 4214
QY 24333 GCTCATTCAGGACATTCGCGGAGGCTTCCGAGGCTTCCGAGGCTTGGAGGAGGAGGAGGAG 24392
Db 4215 CTTCTCGGCTTCTGCGGAGGCGGCGGCGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAG 4274
QY 24393 GACGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 24452
Db 4275 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4311
QY 24453 ACTGAGCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 24512
Db 4312 -CTACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4370


```
Db 11534 GTACCCGCGGGGTCTTGGGACGCTGTCTCTGATGAGGCTATGGAGGAGGCGGTG 11593
QY CCAACCCCTCTTGGTACGCAACACCAAGCGCACACACACACACACACACACACACACACAC 23352
Db 11594 GAGGCTCGGCTGTGTGTGCTGTCTCGGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCG 11653
QY ACACACCCCAACCAAGCCCAAACTGGGAGCTCGCGCGCACACACACCTCTCTGAAACACCC 23412
Db 11654 GTCCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11710
QY ACCACACCGCGGATCATGACCTCCACACACACACACACACACACACACACACACACACAC 23472
Db 11711 ACCCGGTGGGCGGTCTCTGTGACCTCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 11770
QY CTACCCAAACCTCTACCCAAACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 23532
Db 11771 GCGGTGAAACGCTCGCGCGTCCGAGGACCATGCGCGTGGCGCGTGGCGCGTGGCGCGTGG 11830
QY ACCACACCGCGCGCTACCCCAACACACCTCACCCCAACCAACCAACCAACCAACCAACCAAC 23592
Db 11831 GCGCGCGCTCACAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11890
QY ACCCCCAACCAACCAACCTCATACCGCGGAAACCGGCGCGCGCGCGCGCGCGCGCGCGCG 23652
Db 11891 CGCGCGCGCGGACGCGTGTCTGTACCGGTGGCAACCGGCGCGCGCGCGCGCGCGCGCGCG 11950
QY CACCACTCACCAACCAACCAACCAACCAACCAACCTCTCTCACAGCGCGCAACCGCGCGCG 23712
Db 11951 CGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12007
QY CACACCCCGCGCGCAACCAACCTCACCCCAACCAACCAACCAACCAACCAACCAACCAACCA 23772
Db 12008 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12067
QY ATCACCCTGCGACACCAACCAACCAACCAACCAACCAACCTCTCTCAACCAACCAACCAAC 23832
Db 12068 TTCCGCGCTGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12127
QY CCACACACCCCTCACCAACCGCTCATCCACCGAGGCGTCAATCTCTCGCGCGCGCGCGCG 23892
Db 12128 GAGGACACCCGCTCACCGCGGTGTTCCACCGCGCGCGGTACCGGAGTCAACCGCGGTG 12187
QY TCGGAAACCGATCGCGCAATCTCTCTCTCGGTACCGGAGCGCAACCGCGCGCGCGCGCG 23952
Db 12188 TCCGCTACCTCGCGCGCGCGCTGCGCGAGTGTACCGCGGCAACCGCGCGCGCGCGCG 12247
QY ATTCTGATGAGTGTCTGTGACCAATGAAC---GCTTGAACATTCATCTCTCTCTCTCT 24009
Db 12248 CACTGGAGCACTGACCGCGCAACTCGCGCGCGGACTCGACGCGTTCGTCTCTACTCC 12307
QY TCGGCGCGCGCGCTTGGGCGAGCGGGAATCAGTGGCATCTCGCGCGCGCAACCGCATAC 24069
Db 12308 TCGGCGCGCGCGCTTGGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12367
QY CTGAGCGCGCTCGGAGCATGCTCAGACACATGAGCTTCCCGGCGCATCGATCGCGTGG 24129
Db 12368 CTGAGCGCGCTCGCGCGCGCGGTGCGGCGGAGCGACTCCCGCGCACTCTCATCTGCT 12427
QY GCGCCCTGGCGCGGAAAGGCGATGTGCGCGGTGATGCGCGCTCATGTTTACTTGAAG 24189
Db 12428 GCGGTGTGGGCGCGCGGTATGGGCGCGGCGCGGCGCGGCGCGGCGGCGGATCTGCGCGCG 12487
QY CCGCGGATTCGCGATGAGCCACGATGCGCGCTCGCGCATTCATCGTGGCGGCGG 24249
Db 12488 CCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12547
QY CAGCGCGCGAATTCGAACCTGATCATCGCGGACATCGACTGGGAGCGCTTCGTCCCGCG 24309
Db 12548 TCCGG---GGAACCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12604
QY TTACCGCTGCGACGCGCGCGCTCATCGAGGACATTCGCGAGGTTTCGCGAAGGCTTCG 24369
```

```
Db 12605 TTACCGCTTCCGCGCGCGCGCGCGCGCTCTGATCGCGG-----GGCTCGGACGCGCGG 12654
QY CAGGAGCTGGAAGCAGCTGCTGCTGCGCGCAAGAGACGACACAGCTCAGCGGATTCGCGAG 24370
Db 12655 CGCGCGCGG-----GGCGGCGGAGACCCCGAGGAGGGAACGCCACC 12697
QY TCTCTCCGTAGCGATTTGCGCGAGTGCCTCTCAAGCAAGAACCAAGGTGTCTCTCGC 24430
Db 12698 GCTGGCGCGGAC---TCACCGCGCTGCGCGCGCGCGAATCCGACCGCTGCGCGAG 12754
QY CTGATTCGAGCAGGATCTGACCGCTTCTGCGCGCTTCTGTAATCCGGAAGGATTCGAGGAC 24490
Db 12755 CTGCTGCGAGCGCGACCGCGCGCTGCGCGCTCGACGACCGCGCGGAGTTCGCGGAG 12814
QY CAACGAGCTTCCGCGACCTCGCTTTCGACTCGCTGAGTCTGCTGCTGCTGCTGCTGCTGCT 24550
Db 12815 GCGGAAACGTTCCCGCGCATGGCTTCGACTCCCTGGCCACCGTACGGTTCGCGCGGA 12874
QY CTCGCAAGGAAACCGGACTGCGACTCCCGCGCTCCCTGGTCTTGGACTATCCCAACCGCG 24610
Db 12875 CTCGCTCGGCGACGCGCTCGACTGCGCGCGCGATCTCTTCGACCGGACACCGCG 12934
QY CAGGAATGTCTGCGCATCTGCGCACACAACTGCTGCGACTTAGACGACGAAGAGACGCG 24670
Db 12935 GCGCGCTGCGCGCGCGCGCTGCGCGAATGCTTGGCGACCGGACCGGACCGCGCGCG 12994
QY GCACGTGCGAATGCTCTCCCGC 24751
Db 12995 GCGCGCGGACCGGTGCGCGCG 13016
```

RESULT 15

```
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32
```

```
Query Match 7.0%; Score 2140.6; DB 9; Length 11220;
Best Local Similarity 54.0%; Pred. No. 0;
Matches 5775; Conservative 0; Mismatches 4324; Indels 600; Gaps 44;
```

```
QY 19950 CCGGAGTTCCAGCCCGCAAACTCACTCCGCTTCCGCGACCGCGGCGAGCTCGATGA 20009
Db 48 CACGCGGACCTCCACAGGCGCGGTGGCGCGCTCGCGAGCTGGAGGCGAAAGCGGCGGA 107
QY 20010 GCCGATCCCATCGTTGGCATGGCTGCTCCCGCGGCGAGTGAACCTCGGCGGACGA 20069
Db 108 GCGGTGGCATGCTGCGCATGGCTGCGCGCTGCGCGGCGGCGGTGCGCTCGCCCGAGA 167
QY 20070 CTTCTGGATCTGATCTCTCTCGGACGAGCGCATCGCGCGATTCCTCCACCGACCGCGG 20129
Db 168 CTTGTGGCGCTGTTGGCGCGGCGGAGGACGCGATCTCGGATTTCCCGCAGACCGCGG 227
QY 20130 CTGGGACCTGAGACGCTCTTAGACCCCGGACCCCGACCAACCCCGGACACTGCTACACCG 20189
Db 228 CTGGGACCTGGAGGCGCTGTACGACCCCGGAAACCCGAGGCGCACCGGCAAGAGTTACGCGCG 287
```


Db 7364 GAGAGCTCTGGAGCGTCTCTCGCCCGCCACCTCTGTGAGCGTACCGGGGTGCGCCGGCTG 7423
QY 18973 CTCCTACACGCGAAACCGGCCCCCACCACCCCGCCACGCAACAACCTCTACACCCCACTC 19032
Db 7424 CTGCTGGTAGCGCGCGGGGGCGAGCGCCCCCGCGCGCGCCGACCTTGGGCGAGGACCTC 7483
QY 19033 CAACAAAAGGCAATCCACTCTACCATCACCACTTGGACACCAAGCAACCCAGACCAACTC 19092
Db 7484 GCGGGCTCTGGCGGAGGTGGCTTCGCGCGCGCGAGCGCCCGCGAGAGCTG 7543
QY 19093 CAACAACCTCTAAACAGCATCCCGCCCAACAACCCCTCTACACCGTCTATCTACACCGCA 19152
Db 7544 GCGCGGGCGATCCCACTCGTCCCGCGCGAGCATCTGCTGACGGCGCGCTGCTGCAACGCGG 7603
QY 19153 GGCATCTCGACAGCGCACCTCTACCACTCACCACTTCAACCCCACTCACTCAACAGTCTCTC 19212
Db 7604 GGAGTCTGACACGCGAGCGTGGAGGCGCTCACACCGGAAACGGCTGGACGCGGTACTG 7663
QY 19213 CGCGCCAAAGCCACAGCGCCACCTCTCTCCAACTCTACCCCAACAACAACCCCTCAAC 19272
Db 7664 CGCCGAGGTCTGACCGCGGTGGAACCTTGACAGCTCACCAAGGACCTTGGCTGAC 7723
QY 19273 GCCTTCTCTCTCTCTCTCGCGCGCGCACCTTCTGGGGCAACCGCGCAAGCAACTAC 19332
Db 7724 GCCTTCTCTCTCTCTCTCTCGCATCTCGGCATCTGCGGCAACCGCGCGCGGCAACTAC 7783
QY 19333 GCGGAGCAAAACCTCTGAGCGCTCGCCACCGCCACCGCCACCACTCACTCACTCTCC 19392
Db 7784 GCGGCGGCAACAACGCGGCTCTGACGCGCTCGCGCCACCGCGCCCAAGCGGCTTGGCC 7843
QY 19393 GCCACAGCATCCCTTGGGGCACTTGGCAAGGAAACGAGCTGGCGACTGTCAAGTCAG 19452
Db 7844 GCCAGTCTGCTGCTGGGCTCTGGAGCGGACACGCGCATGGGCGGCGCATCTGGC 7903
QY 19453 GAACATCTCGCGCGCGCGGAT-----GTTGCGCATGCGCGCGAGTTGGCGGTC 19503
Db 7904 GCGCGCGACCTCGCGCGCTGGAGCGGCGCGGCGGAATCACCCGCTCACCCGCTGCGAGGCG 7963
QY 19504 ACAGCTGTTGAGCGCGGATCGGAGGGGCGCGGAGTCTCTCTGTCGCGCATTCGAC 19563
Db 7964 CTGCGGCTCTTACAGCGCGGCTGCGGAGGAGCGCGCTCTCTGTCGTCGCGCGGCTC 8023
QY 19564 TGAAGAAATGGGACCGGTTCTCTCAGCAAGTCTGCTGCTTGTCTGAGGACCTTCC 19623
Db 8024 GGTCCACCGCCACCGGGGCGAGCGGACGCTCTCTGCTGTGGCGGCTCTGTC 8083
QY 19624 CAGGCAAGGAACTGAGGAGCGCGGAGTACGTTGAGCAGCGAGAGCAACAACCTC 19683
Db 8084 CGGGCGCGCGCGCGCTGCGCGGGAACGCGCGCGGAGCGCGGCGGACACGACCGCGGCG 8143
QY 19684 CGGCAACTC---CTCATGGTCTGCTACGTTCCGAGCAGGAGAGAGTGTCTACGCTC 19740
Db 8144 TGCTGAGCGGCTCGCCCGACAGTCTCCCGAGGAGCGCGGAGCAAGCTGTCAGCTC 8203
QY 19741 GTCCGCACTCACCTCGCGGAGTCTGCGCGGAGAGTCTCGAGGCACTCCCGCGGCT 19800
Db 8204 GTGACGGGTGTCTGTCGAGAGCTCTCGGGGACCGCGACTCTCGCGGCTCGGGGCGGAG 8263
QY 19801 CGGCTGTTCAGGAACTAGGGTTGACTCGCTTGGCGCGGTGGAGTCTCGCAACCACTC 19860
Db 8264 CGGCTCTTCAAGGACCTCGGCTTCGACTCCCTGGCGGGGTGGAGCTCGCAACCGGCTG 8323
QY 19861 GAGCAGACAGGAGTGTGCTGCGGAGTCTGCGAGCTCTGCTCTGATTACCCAGCCGCTC 19920
Db 8324 AAGCGCGCACCGGCTCGGCTTCCCGGAGCAGCGTCTTTCGACCATCTCTCGCGGCG 8383
QY 19921 AAGCTCGCCAAATTTCTGCTCTCCGAGATCGGGAGTTCCAGCCCGCAACTCAACTCG 19980
Db 8384 GCGCTCGCTCCCATCTCTCGCCAGGTGCGCGGTTGAGAGGGGAGCGGGGAC 8443
QY 19981 CTTTC-----GCGACCCCGGCGAGAGTCTGATGAGCG 20013

Db 8444 GCGACCGTCTGTGGCGAGCGGGCGCTTCTCTGCTGACCGTGGACGACGACGATCG 8503
QY 20014 ATCGCCATCTGTGGATGCGCTGTGCTTCCCGCGGAGTACCTCTGCGGCGAGCACTTC 20073
Db 8504 ATCGCGATCTGTGGCATGCGATGCGCTATCCGGGTGGTGTGCTGCGCGGAGGACCTG 8563
QY 20074 TGGGATCTGATCTCTTCCAGCAGGACGCGATCGGCGGATCTCCCAACCGACCGGCTG 20133
Db 8564 TGGCGGCTGTGGCGGAGGAGCGGATCAGCGAGTTCCTCCGCTCAACCGCGCTG 8623
QY 20134 GACCTGGACACGCTCTACGACCCCGACCGGACCGGCGGACCTGCTACACCCGAAAC 20193
Db 8624 GACCTGGAGAGCTCTTACGACCCCGATCCCGAGTCCGAGGCGACCACTGCTGCGCGGAG 8683
QY 20194 GCGGATCTCTTACGACGAGCGGCTTCTGAGCGCGGATCTTGGGATCAGCGCCCGC 20253
Db 8684 GCGGGTCTCTGGAAGGCGCGTGTGACTTTCGAGCGCGCTTCTTGGGATCTGCGCGCGC 8743
QY 20254 GAAGCCCTCGCCATGAGACCCCGACGACGATCTCTCTCGAAACCGCTTGGGAAACCATC 20313
Db 8744 GAGGCCCTGTGATGAGACCGCGAGCGGCTGTCTGAGAGTGTCTTGGAGGCGCTG 8803
QY 20314 GAAAGCGCGGATCAACCCCGACCTCTCAAGGACCGGCGGACCGGAGTCTTACCGCGC 20373
Db 8804 GAACGCGCGGATCGACCGCTCTGCTGCGCGGCGGCGGCTGTGCTACGCTGGC 8863
QY 20374 ACCAAGCGGACGACGCGGCGACATCTGCTGAGGCGCGGAGGATCAGGAGGATC 20433
Db 8864 GCGGCGAGCGGCTGTGATGCGCTTCCGATCCCGGCTGTGCTGCGGAGGCTTGGAGGCTAT 8923
QY 20434 GTCTGACCGGGGCGGACCAAGCATCTGCTCGGCGGAACTCTCTTACATCTCTCGGTTG 20493
Db 8924 CTGTGACCGGCGGCGGCGGCGGCTGTCTCGGCGGATCTCTACGCGCTCGGCTC 8983
QY 20494 GAAGGCGCTGCGGTCACCTCTGACACAGGCTTCTCTCTGCTGCTGCGCTTACGCTC 20553
Db 8984 GAAGGACCGTCTCATGACGCTGAGAGCGGCTCTCTCTGCTGCTGCTGCTGCTGCTG 9043
QY 20554 GCTGCGAGTCTCTCAGGTCGCGTGAATGACCATGCGCTTGGCGGCGGCGGCGGCGGCT 20613
Db 9044 GCGTACGCGGCTCTGCGGCGGCGGAGTCTGCGGCTGCGGCGGCGGCGGCTGCGGCTG 9103
QY 20614 ATGACACCGCGGATCACTTCAACGAAATTCGCGCGGCGGAGTCTGCGCGCGGCGGCT 20673
Db 9104 ATGCGGATCGGCGGCTTCTGAGGATCTCTCCGCGAGAGGCGTGGCGCGGCGGCG 9163
QY 20674 CGTTGCAAGCGGCTTCTGCGGCGGCTGACGCTGACGCTGCGGCTGAGGCTTGGGATG 20733
Db 9164 CGCTGCAAGGCGTCTCTGCGCGCGGCGGCGGCGGCTGCGGCGGCGGCTGCGGCTG 9223
QY 20734 CTGCTGTTGAGGCGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCTG 20793
Db 9224 CTGCTGTTGAGGCGCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCTG 9283
QY 20794 CGTGGAGTGGCGTCAACAGGAGCGGTCGAGCAACGCTTCTGACCGCGGCGGCGGCGGCGG 20853
Db 9284 ACCGCGACCGGCTCAACAGGAGCGGCTCTCAACAGGCGTGAACGCGGCGGCGGCGGCG 9343
QY 20854 TCCGAGGCGGCTCATCGGCGGCGGCTTCCGCAACCGGCGGCTGACCGCGGCGGCGGCTG 20913
Db 9344 GCGGAGCAACGCTCATCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCTGCGGCGG 9403
QY 20914 GATGCGTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 20973
Db 9404 GACGCGTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9463
QY 20974 ATCTTCGCGAGCTTACGAGAGCGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 21033
Db 9464 CTGCTCGCGCTCTCGGAGGAGCGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 9523
QY 21034 AAGTCAACGCTCGGAGACACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21093
Db 9524 AAGTCAACATCGGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 9583

Qy	17413	ATCGCGCCATCAACACACCCCACTCCTCGTGCATCAGGGGACACCCCAACACCGTCCAA	17472
Db	5204	GTCCGCGCCGTCAACGGACCAACCGCGTGGTCTCTCGCGACCGGACCGCGCGAG	5263
Qy	17473	CACATCACCACTCTGCGCAACACAGGATCAAAACCAAAACCTTCCCAACCAAAAC	17532
Db	5264	GAGATCGCGCCGTATGGGGGACCGCGCGGACCGCGAGGTGCGCGTCAAGCAC	5323
Qy	17533	GCCTTCACCTCCCGCACACCAACCCCATCTCAACCAACCTCAACAGCACACCAAC	17592
Db	5324	GCCTTCACCTCCCGCACATGGAGAGGTCTCGACGAGTTCTCTCGGGTCCGCGAGGC	5383
Qy	17593	CTCACCTACACCCACACCCCTCATCACCGCCAAAC-----CCCA	17640
Db	5384	CTGACCTTCAGGAGCGCGGATCCCGTCTCTCAACGCTCACCGCGCGCTCGTCAAG	5443
Qy	17641	CCGACCACTCTCAACCCCACTACTGACCAACAGCCGCAACACCGTTCGACTAC	17700
Db	5444	TCCGGGAGTCACTCGCCCGGTACTGGGTGACAGATCGCGGGCGGTGCGCTTC	5503
Qy	17701	GCACCAACCAACCAACCTTCAACCAACCGGCTCACACCTTACATCAAACTCGGACCC	17760
Db	5504	CTGAGCGCGTCCGACCTTGGCGGCCAGACGCGACGCTCTCTCGTCAAGATCGGCC	5563
Qy	17761	GACAAACCTTCAACCACTTCAACCAACCACTTCCCAACACCCCAACCAAC-----	17816
Db	5564	GAGCGCTCTCAACGCACTCGCGAGGAGCTCTCGCGCCGCGCACGACGCGCCCGGAC	5623
Qy	17817	-----CCTCACCTTCAACCAACCCCAACCAACCCCAACCACTCTCTCAACCAAC	17868
Db	5624	GCCCGGACGTCAAGTCTCGCTGCTGCGCGCGGGCGCCCGAGCCGAGACCTC	5683
Qy	17869	CTCGCAAAACCAACCACTGGCACCCCAACCACTACACCAACCAACCAACCAACCC	17928
Db	5684	GCGCGCGTCTCGGACCGCCATGTCCAGCGGACCTTGGACCGGGGCTGTTCTTC	5743
Qy	17929	CACACCAACCACTTCAACCTTCCCACTTCCCACTTCCCAACCAACCAACCACTTGGCTC	17988
Db	5744	CCGACGGGCGCGCACGCACTTCCCACTTCCCACTTCCCAACCAACCAACCACTTGGCTC	5803
Qy	17989	GAAGACACACGCGGCTGCGGATCCGTTCCGTTCCGTTCCGTTCCGTTCCGTTCCGTTCC	18048
Db	5804	ACGCGGAGCGCGTACGGACCGCGGCTTCCGTTCCGTTCCGTTCCGTTCCGTTCCGTTCC	5863
Qy	18049	CGCGCGGACGCGAGGTGGAGTCCGCTTCTGGGACCGCGTGGCGCGGACCGCTG	18108
Db	5864	CTGACGACACGTTGAGGTTCGCGGCGGACGCGGCTCTGCTGACCGCGCTCTCTCC	5923
Qy	18109	GAAGCGTCCGACCACTGCGCTGCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCCCG	18168
Db	5924	CTGACGACCACTGCGCTGCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCC	5983
Qy	18169	GCATCTCC-----	18177
Db	5984	ACGCGCTTCTGGAGTTCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCCCGCTTCCCG	6043
Qy	18178	-----GCTGGCACCGCCACCAACAGGAGGAGTTCGCGGCTGCGGAG	6043
Db	6044	GAATCACCTCGAAGCGCGCTGCTCTCGCGGCGGACCAACGCTGCGGCGGCTGCGGAG	18207
Qy	18208	CGCATCAACCTTGGACCTTACAGGAACCTTGAACCTTCAACCTTCCCGCTTCCCGCTTCC	18267
Db	6104	GGCTGAGCGGACGCGAGTCCGCGGCGGCGGACCTTTCGTTGTGTACAGCACACCC	6163
Qy	18268	CAACCCCAACAACTGGCTCATGCGCATCCCGAA-----	18303
Db	6164	GACTCGGCGACACCGGTGACGACGCGCGCGCGGAGTGGACCGCGCATGTCTCCGCGTA	6223
Qy	18304	-----	18303
Db	6224	CTCGGCAAGGGGACCCCGGCAACGAGTTCGGAACACCCCGGACCGACGCGGAGGTTTCA	6283

Db 3080 ACCCGCGGACACACCGCGGACCCATCGCATCATCGGCATGGCCTCGCCGTACCC 3139
Qy 15256 GGTGATGATCGGACCGGTGATGATCTCTGCGAGGTGGTGTAGTGGTCCATGACGGATC 15315
Db 3140 GCGGCGTCACTTCCCGGAGGACCTGTGGCGGTCTGTCCCGAGGGGCGGACGCCGTC 3199
Qy 15316 GCGGATTCGCGAACCGTGGGTGGGACCTCGACACGTGTACACCCCGGACCCGGAC 15375
Db 3200 TCGGGCTGCCACCGACCGCGGTGGGACGAGGAC---CTCTCGAGCGCGACCCGAC 3256
Qy 15376 CACACGGAACAGCTACACCGGAGCGGCGGATTCCTTTTACGACGAGGCAATTCGAT 15435
Db 3257 CGACGCGCAAGAGCTCGGTCCGCGAGGCGGATTCCTGACACGCGCGCTGTTCGAC 3316
Qy 15436 CCGGACTTCTTCGATATCATCTCGCGGTGAGGCACTGGCGATGACCGCGAGCAGCGCTG 15495
Db 3317 GCGGCTTCTTCGGGATATCGCCCGGAGGCGCTCGGCATGACCGCAGCAGCGGCTG 3376
Qy 15496 CTGCTGAAACAGCGTGGAGAGCATCGAACAGCGCTGCAATCAACCCCGACACGCTCCGT 15555
Db 3377 CTCTGGAGACGCATGGAGGCGGTGGAGCGCGAGGCTCGACCCCGAGGCTCAAG 3436
Qy 15556 GGCACACCAACCGCGGTTCGCGGCGGTGACCTTACACGACTACGCGCGCGCTTCC 15615
Db 3437 GGCAGCGGACGCGCGCTTCTGTCGCGCGCACCGCTTGAATACGCGCGCGCATGCAC 3496
Qy 15616 ACAGCTCCGCGAGGTTGAGGGGTATCTCGGCGAGGACCGCAGGCGATATCGCTCG 15675
Db 3497 GACGGCGCGAGGCGGTGAGGGGCACTCTACCGGGACCAACCGCGAGTGTGCG 3556
Qy 15676 GGTGCTGTCCCTACGCTCTCGGCTGGAAGTTCGCGCTCTACAGTCGACACTGCTCG 15735
Db 3557 GCGCGCATCGCTACAGCTCGGCTCACCGTCTCTGCGGTACCGTGCACGCGCTGC 3616
Qy 15736 TCTTGTCTGCTGTGCTGTGCACTTGTGCGCTGTGAGCGCTGCGGTGCGCGAGTGTTC 15795
Db 3617 TCGTCTCTGCTGCTGCGCTGCACTTGTGCGCTGCGTGTGCGCGAGGCGAGTGCAGC 3676
Qy 15796 ATGGCTCTCGGCTGCGCTGACGCTGATGTCAACCCCGCGGCTGTGAGTTCG 15855
Db 3677 CTGCGCTCGCGCGGAGGCGACCGTATGTCAACACCGGCGATGTGCTGAGTTCG 3736
Qy 15856 CCGCAGCGGCGCTGCGCGTGAACGCGCGGTCAAGCGGTTCGCGCAGCGGTGACGCG 15915
Db 3737 CGCAGCGCGCTCGCGCGGACGCGCGCTCAAGGCTTCTCCACTCGCGCGAGCG 3796
Qy 15916 ACCGCTGGGTGAGGTGTGGAATGCTGTGAGCGGTGTGAGCGGTGTGAGCGCGCG 15975
Db 3797 ACTCTGCGCGGAGGCGTGGGCTCTGCTGTGAGCGGTCTCGAGCGCGCTCGAGCGC 3856
Qy 15976 CTGCGTCAACGATCTCGGCTGCTGCGGTGAGTGCAGTGCAGTGCAGTGCAGC 16035
Db 3857 AACGGCACCCCTGCTCGCGGTGATCCGGGGAGCGCGGTCAACAGGACGCGCTCC 3916
Qy 16036 AACGGCTGACGCGCGCCCAACGCGCGTCCAGAGCGTGTATCGCGCTGCGCTGCGC 16095
Db 3917 AACGGCTACCGCGCGCGCGCGCGTCCAGCAGCGGTATCCAGAGGCGCTGCGC 3976
Qy 16096 AACCGGACCTGACCCCGCGCGAGTGTGATGAGGTGAGGCGCGACCGGCACT 16155
Db 3977 GACCGCGGCTACCCCGCGCGAGCTCGAGCGCGTTCGAGCGCGGTACGCGTACCGG 4036
Qy 16156 TTGGCGACCGATCAGGCGCGCGCTCTCGCCACCTACGAGACGACCGCGCGCG 16215
Db 4037 CTCGGGACCCCATCAGGCGCGAGGCGATCTCTGCGCACCTACGCGCGGACCGGCG 4096
Qy 16216 AACGAACCGTGTGGCTGAGTGAAGTGAACATCGGCACGCGAGGTGCGCGCA 16275
Db 4097 GCGCTCCGCTCCAGTCCGCTCGTGAAGTGAACATCGGCACGCGAGCGCGCGCG 4156
Qy 16276 GGTGCGGCGGTATCAAGATGATGGCGCTCGGAAATGTTGCTGCGCGGAGC 16335
Db 4157 GCGTGGCGGCTCATCAAGATGTTCTCGGATGCGCCACGCGCTCTGCGCCAGGACG 4216
Qy 16336 TTGCAATGATGAGCGCTCGCGATGTGGATGTGTCTCGCGGCGCGTGTGAGCTGCTG 16395
Db 4217 CTCACGTGAGCGGCGCACCAACCGCTGACTGGAGCGCGCGCGGTTCGAGCTCTCT 4276
Qy 16396 ACAGGACCGTTCCTGCGCGCGGAGGCGCGCTCGCGGCGGAGGAGTGTCTATCG 16455
Db 4277 ACCGAGGCGGAGTGGCGGCA---GACGGCGCGCGCGCGCGCGCTCTCTCTC 4333
Qy 16456 TTGCGGTGAGCGGCGCACCAACCGCTGATCTCTGGAAGAGCACCGCGCCACAACATC 16515
Db 4334 TTGCGCATGAGGCGCACCAACCGCTGATCTGTTGTAACAGCGCGCGCGGAG 4393
Qy 16516 CCGTACAGACACCGCGCGAGCGCGCGGAGAGCGCGCGCGATGTTTCGCGG 16575
Db 4394 GCGCGGTCAACACACCGCGCGGAGAGCGGAG---CGGG 4435
Qy 16576 GAAGCGCGCGGACGACCGCGGTACCGCGGCGGAGAGCTGCTCTGTCGCGGACGT 16635
Db 4436 GAAGCGCGGACACACCGCGCACCGACCGCGCGCGGTGCG---CGTCCCGGAA 4489
Qy 16636 CAGGCGTGTGCGCGGTGTGTGCGCAAGTGTGCGCGCGCGCTGCGCGCGCGAGGCG 16695
Db 4490 CCGGTACGCGCGCGGTGTGTGCTCTCGCGCGGAGCGCGCGCGCTGCGCGCGCGGCG 4549
Qy 16696 CAGGCGTGCACGCGCGCTACCGCGACCGCGCGCTCGACTCGCGCGCGTTCGCGTAC 16755
Db 4550 GTTCGCTGCGGACCTTCTCGACGCGCGCGAGCGTCACTGCGCGCGCTTCGAGCGC 4609
Qy 16756 ACCCTCGCGCGCGCGCGCTGTTGACACCGCGCGCGCTCATCGCGCGCGACCGC 16815
Db 4610 TCGCTGCG---CGCGGTGACCGCTTTCGAGCACAGCGCGCTTACACCGCGCGCG 4666
Qy 16816 GACACTTCTGCAAGACTTCAGGACTGCGCGAGGCGAAGCGCGCGCGCTGATC 16875
Db 4667 GACGAGTGTCTCGCGCGCTCGACCGCTTCGCGCGCGGAGCAAGCGCGCGCTGCTC 4726
Qy 16876 CACAGCGCGCGCGGCGGACCGCGCGCGGAGCGCGCGCGCGCGCGCGCTTCCTC 16935
Db 4727 -----ACCGCGGAAACCGCGCGCGCGCGCGCGCGCGCTTCCTC 4765
Qy 16936 TGCTCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACACACCGCGCGCG 16995
Db 4766 TTCACGCGCGGAGCGCGCGCTGCGCATGCGGAGGAGTTCGCGCGCGCGCGCGCG 4825
Qy 16996 GTCTTCG 17055
Db 4826 GTGTTGCG 4885
Qy 17056 CTCTCTCCCTCTCACCGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17115
Db 4886 CTGCGGAGATGCTCGCG-----CGCGCGCGCGCGCGCGCGCGCGCGCGCG 4906
Qy 17116 GCGCGACTGCTTCAGCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17175
Db 4907 GGGGAGGAGTGGACCTCACCGGTACACCGCGCGCGCGCGCGCGCGCGCGCGCG 4966
Qy 17176 TTCACCGCGCTCTCACCGAGGCTTACCAATCACCGCGCGCGCGCGCGCGCGCG 17235
Db 4967 GTGTTCCGCTCTCT---CGAACACACCGCGCGCTGCTCGCGCGCGCGCGCGCG 5023
Qy 17236 CTCGCGGAAATCACCG 17295
Db 5024 GTCGCGGAGATGCG 5083
Qy 17296 CTCATCACCG 17352
Db 5084 CTCGCTACCG 5143
Qy 17353 CTCACACACCG 17412
Db 5144 GTGAGGCG 5203

13096 GGAACCTTCAAGTCCAAACATCGGGCACTCCATGGCGCTGCGGGTGTGGCGGGGTCTATC 13155
1037 GGTCTGGCCAAAGACGAACTCGGGCACTCGAAGGGCGCGCGGCAATCGTCGGCTCTATC 1096
13156 AAGATGGTATGGCGTGGGAAATGTCTGCTGCCCGGAGAGTGTGATGTGATGAGCG 13215
1097 AAGAGCTCTCTGGCGTGGCGCGCGGATCCCGGCGAGCTCAACTTCGTAGCGCC 1156
13216 TGGCCGATGTGACTGTGTCTGGGGGTGGGTGAGTCTGCTACGAGAGCGGTGCGCTGG 13275
1157 CACCCGGACATCCGCTGACACCTTCGGGCTCGAGCTGCCGACGCGCTGGGGAGTGG 1216
13276 CCGGGGGGAGGGGGGTACGGCGGAGAGTGTCTATCATTCGGCGCTCAGCGGACCC 13335
1217 CCGACCCGAC---CGGAACTCTCGCGGGTCTGCTGTTGGCATGGGGGACCC 1273
13336 AAGCCACAGTCTATCTCGAGGAGACCCGGCCCAACATCCCGTCTAGACACACCGCC 13395
1274 AAGCCACAGTCTCTAGCGAAGGCCCGCGCCAG----- 1309
13396 GACGACGTCCGGGAGAAATCAGCGCGGAGAGGATGCCGCTAGTGGCGATGAGGTGTCT 13455
1310 -----GGCGGCGAGAGCCCGGCATCGATGAGGAGACCCCC 1345
13456 GCGGCGAGTCCAGGGGTGGCGGTGGCTGGTGTGGCGCAAGTCCGAGCGGCGCTTGGC 13515
1346 GTGACAGAGGGGCGGACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1405
13516 GCCAGGCGGAGGCGCTGACGCGCCACCTCACGACCAACCGCGGCTCGACCTCGCGGAC 13575
1406 GCCAGGCGGCGGCTGACGAGGCGCTCGAAGGAGGACCCGAGGCTCGCGCGCGCGCA 1465
13576 GTCCGCTACACCTCGGCGCAAGCGCGGGCGGTGTTGACAGCAGCGCGGCACTATCGGC 13635
1466 CTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1525
13636 GCGGACCGGACACTTCTGCAAGCACTCCAGGCACTGCGCGCAGCGCAACCCACCGCC 13695
1526 CCGGACCGGCGGCTCTCTGACGCGCTCGGCGGCTGCGCGCGGAGCGCGCGCGCC 1585
13696 GCGGTATCACAGCAGCGCGCCAGGCGGAGCCGCGGACCGGGAGCGCGGAGGAGGAAAGCC 13755
1586 GCGGTGCTACCGGCAACCC-----GCCCGCGGCGCCTC 1621
13756 GCATTATCTGCTCGGACAGGCGACCCAGCGCGCGGCAATGCGCCACGCGCTCTACAC 13815
1622 GCGGTCTGTTACGCGGCGAGGCTGCGCAACGTTACGCGGCAATGGGCAATGAGTTGTAGCC 1681
13816 ACCACCGGCTTTCGCGCGGCACTCAACGACATCTGACCCACCTCGACCCCGGACCTC 13875
1682 GCCACCGCGCTTTCGAGCGGCTTTCGAGCGCGTTCGCGCGGCACTGACCCCGCTCTC 1741
13876 GACACCGCTCTCTCCCGCTCTCTCACCCAGGACCCCAACACCCAGGACACCAACCCCTC 13935
1742 GACCGGCGCT-----GGCGGAATC 1762
13936 GAAGAGCGCGGCACTGCTTCAGAGACACCGCTAGCGCGGCGCGGCTCTTTCGCGCTTC 13995
1763 GTCCGCGCGGCGGACACCTCTGACCGGCACTGCTTCCACACAGCGCGGCTCTTTCGCGCTG 1822
13996 CAGGTGCGCTCTCACCGCTCTCTCACCGGCTTACACATCAACCCCGGCTACTACGCG 14055
1823 GAGGTGCGCTCTCACCGCTCTCT---CGAGTCTGGGGGTCTACGCGGCTGCTGCGC 1879
14056 GGAACCTCTCTCGCGGAAATCACCGCGGCGGCTCTGCGGCGGCTCTCTACCTCTACCGGAC 14115
1880 GGCACCTCTCTCGCGGAGATCAGCGCGGCGGCTCTGCGCGGCTCTGCTGCTGCGGAC 1939
14116 GCCACACCTCTATCACCGGCGGCGGCTCTATGCAACCAATGCGC-----CCCGGACCC 14172
1940 GCGCGCGGCTCTGTCGCGGCGCGCGGCTCTATGACGCGGCTCTCCCGGAGCGCGCG 1999
14173 ATGACCACTCCACACACACACACACATCACCACTCACCGCCCGGAAAC 14232
2000 ATGTCGGGTTCGAGGCGAGCGAGGAAGTCTTCGCACTTCGCGGAGCGGAGCGG 2059
14233 GACTTCGCACTGCGCGCATCAACACCCCACTCTCTCTGTCATCAGCGGACCCCGCAC 14292
2060 GAGTCTCTCTTCGCGCGGTGAAGCGCGCGCGGCTCTCTCTGCGCGGCGGAGCGC 2119
14293 ACCGTCCAAACATCACCACTCTGCGCAACAAAGGCAATCAAAACCAAAACCTCTCCC 14352
2120 GCGGTCTTCGAGCTGCGCGAGTCTGCGCGAACAAGGCGCGCGGACGAGGGGTGAGC 2179
14353 ACCAACACGCTCTCACTCCCGCCACACCAACCCCATCTCAACAACTCCACGAGAC 14412
2180 GTCGCGACGCTCTCACTTCGCGCTCATGAGCGGATGCTCGAGACTTCGCGCGGCTC 2239
14413 ACCAAACCTCTACCTACACGCGCCACACCGCTCTCATCCGCTCATCCGCGCAAC 14465
2240 GTCGAAGAGCTGGAATTCAGGAGCGCGCTGCAAGTCTGTCACGGTACCGGCGCTG 2299
14466 -----CCACCGGACCACTCTCACCGCCCACTACTGACCCCAACAGCGCGCAACACC 14520
2300 CTGTTCAGCGGCGCAATGAGCGATCCCGAGTACTGGGTGAGACGAGTTCGCGAGGCC 2359
14521 GTCGACTACGCGACCAACCACTCTCAACCAACAGCGGTCAACCACTTACATCGAA 14580
2360 GTACGCTTCTCGACGCGCTAGCGACCTTGAGGAAATCGGGCGCGACACTTCTCGAG 2419
14581 CTCGAGCCGAGCAACACCTCTCAACCTCAACCAACAGCAACTCTCCCAACACCCCGAC 14640
2420 CTCGCTCCGAGCGGCTGCTGCTGCGATGGCGGCGACTCTCGTACGCGACGAGAGCC 2479
14641 ACCACCTCTACCTCACCGCCACCAACCAACCGCGCGCGCGCGCGCGCGCGCGCGC 14700
2480 GCGACCGGCTCTCGCGCTGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2539
14701 GCGAAACCAACCACTGCGCACCCCGCACCACTACACCAACCAACCAACCAACCAAC 14760
2540 CTCACCACTCTCTGTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2599
14761 ACCACACCACTGCGACTCTCCCGCTACCGCTTTCGAAACCACTTACTGCTGCTCA- 14819
2600 GCGCGGTGCGGCTGCGCTGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2659
14820 ACCACCGCGAGCGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 14879
2660 GCGCGCGCGGAAACCGCGCGCGCTCACGCGGCGCGCGCGCGCGCGCGCGCGCGCGC 2719
14880 ATCAACCGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 14939
2720 GCGCGCGCGCGGCTGC 2779
14940 GTTGGCGCTG-----GTGCGCACTTCTGCGCCT 14967
2780 GGTGGCGGTGATCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2839
14968 GGTGCTGCGGC 15027
2840 GCGCGCTCTGAGTACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 2899
15028 GGTGTTGATTTCTCTCGCGCGCAATTCAGCTTCTGTAATTCAGTCTGCTGCTGCTGCTG 15087
2900 GGTGTTGACT 2959
15088 CCGCTTCTCGGC 15135
2960 CCGCTTCTCGGC 3019
15136 TCGCACTTCTCTCGGC 15195
3020 GCGCGCTGCT 3079
15196 CTAAGTGGCGTCCCGC 15255

QY	21694	GGCATGGCCACGGCCCTTACACACCCACCCCGCTTTCGCGCGCCGACCTCAACGACATC	21753
DB	10151	GGATGGCCCAACACCTGATGCGGCCCATACCTTCTTCGCTGCCCTTCACGAGGTG	10210
QY	21754	TGCACCCACCTCGACCCCACTCGACACCCCTCTCTCCCTCTCTCAACCAACGAC	21813
DB	10211	ACGGACCGTCTGACCCGCTGCTCGCGCGCGCTCGCGCGCTGCT-----G	10258
QY	21814	AAGCAACGACMAACGAGACGGCGCGCACTGCTCCAGCAGACCCCGTAGCCGACGCC	21873
DB	10259	GAGCCCGGACCGCGCTCGCCCGAAGCGGCACTCTTGACCGGACCGAGTACACCGAGCG	10318
QY	21874	GCCCTCTCGCTTCAGGTGCGCTTCCACCGCTCTTCAACGAGGCTACCACTACCC	21933
DB	10319	GCGCTCTTTCGCTCGAGGTGGCGCTTCAACCGCTCTG---GAGCACTGGGGATGCGC	10375
QY	21934	CCCACTACTAGCGCGGACATCCCTCGGCGAATCACCGCGCGCCACCTCGCGCGCATC	21993
DB	10376	CCGACCTGCTGCTGGGCACTCGGTGGGCACTGGCGCGCGCGCGCGCGCTGATG	10435
QY	21994	CTCACCTTACGAGCGGACCACTTATCAACCAAGCGGCGCACTCATGCAACCATG	22053
DB	10436	CTCGATCTCGAGACCGCTGCGCGCTGGTGGCGCGCGCGCGCGCGCGCTGATG	10495
QY	22054	CCCC---CGGCACCATGACACCTTCCACACACCCACACACATCAACCAACCATG	22110
DB	10496	CGCGCGGCGCGGATGCTCTCGTGGCGCGCGGAGGACGAGGTTCGCGCATGCTG	10555
QY	22111	ACCGCCACGAAACGACCTGCGCATGCGCGCATCAACACCGCCACCTCCCTCGTCACT	22170
DB	10556	GCGCGCGGAGGACCGCTGCGTGGCGGTGAAACCGCGCGCGCTGCTGGTGTATC	10615
QY	22171	AGCGGACCCCGACACCGTCAACATCACACCTCTGCGCAACAAAGGCAATCAA	22230
DB	10616	TCCGCGCGGAGGAGCGTGGCGGAGCGCGCGCGCTGCGCGGACGAGCGCGCGC	10675
QY	22231	ACCAAAACCTTCCCAACACGCGCTTCCACTTCCCGCCACACCAACCCCATCTCAAC	22290
DB	10676	ACGAGGGGTTCGCGTGGCGACGCGCTTCCACTACCGGAGACCGGAGTGTGCTGCC	10735
QY	22291	CAACTCAACGACACCCAAACCTTCACTACCAACCCACCCACACCCCTCTCACTACC	22350
DB	10736	GGATTCCGGAGGTTCGCGCGCGCTTGGCTACCGGAGACCGGAGTGTGCTGCTCC	10795
QY	22351	GCACA-----CCCCACCGACCACTCTTCAACCCCTTCACTTGGACCCCA	22398
DB	10796	ACGGTCAACGGGGCGCGCGCGCGCGCTGAACTCACCGCGCGCGCTTCTGGGTGGCC	10855
QY	22399	CAAGCCCGAAACCGTGGACTACGACACCAACCAACCCCTTCCACCAACGCGCTC	22458
DB	10856	CAGGTTCGTGAGCGCTGCGCTTTCGGGACGCGGTTCGACAGGACACCGCTTCGAGGC	10915
QY	22459	ACCCTTACATCGAACTCGGACCGGACCAACACCTTCAACACCTTCAACCAACCTC	22518
DB	10916	CGCACTTCTGAGACCGCGCGCGCTGCTGCGGATGTCAGAGGAGTGCCTG	10975
QY	22519	CCCAACCCCGACACCACTTCACTTCAACCCCGCGCGCGCTTCAACCCCGCAAC	22578
DB	10976	GAGGACGACACCTGCGCTTCTGCGCGGATCCAAAGCCCGGACCGCG-----	11026
QY	22579	CTCTTCAACCACTCGCAAAACCAACCAACCTGCGACCCCGCGCGCGCTTCAAC	22638
DB	11027	-----	11026
QY	22639	CACAAACCCCGACACCCACCTTCAACCTTCAACCTTCAACCTTCAACCTTCAAC	22698
DB	11027	-----	11026
QY	22699	CACCTTGGCTCGAACTACCCAGGCGCGCAACACGCGCGCTCAAGGCTTCTCGCGC	22758
DB	11027	-----CCGACGCTCGCGCGCTTCCGCGCGCTTCCGCGCGCTTCCGCGCGC	11068
QY	22759	TCGGCTTCAGACACCGCGAGTTCGGAGTTCGGGACGCGGTGAACGAGAGACCTCCAG	22818
DB	11069	GCCTGTACGGCGGGGCGCGGTGGACTGGGCGGATGCAACGCGCGCGCGCGG	11128
QY	22819	AGCTTCGGGAAACCTTCGACATCGACGCTTCTGCTTGGACACGGTGGTCCCGCATC	22878
DB	11129	GGGCGCGCGCGCGGTGGAATGCGCTTCACGCTT-----CGGCGAC	11173
QY	22879	TCGGCTTGGCACCGCAACACGACCAACCGCGCATCAACCTTGGACCTTACAGAA	22938
DB	11174	CGCGCTACTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACCGATC	11233
QY	22939	ACCTGGAAACCCCT---CACCTTCCCAACCAACCAACCAACCAACCAACCTTCTATC	22995
DB	11234	GCTGGACCGGCTGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCTTGGTGTG	11293
QY	22996	GCAATCCCGGAAACCGACACCAACCCCACTCAACCAACCTTCTTCAACCTTCCAC	23055
DB	11294	ATCCACCCCGACGACCGCGCTGCGGAGTGTTCGGCCACCGCGAAACGCGCGCTGCGC	11353
QY	23056	CACCAACGCGCATCACCCCATCCCTCTACTGTCAACCAACCAACCAACCAACCAAC	23115
DB	11354	GCGCGCGCGGAGCGCGCTTACCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCGCTG	11413
QY	23116	CTCCACCAACCCCTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	23172
DB	11414	TTGCGCGCACTGCTGCGCTTCCGCGCACCGGACCTTGACACACGAGGTGACACAGC	11473
QY	23173	ATCACCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	23232
DB	11474	GTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	11533
QY	23233	ACACCAACCGCGCATCTTCAACCTTCAACCTTCCCGCAACCAACCAACCAACCAAC	23292
DB	11534	GTACCGCGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	11593
QY	23293	CAACACCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	23352
DB	11594	GAGGCTCGGCTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	11653
QY	23353	ACACACCCACCGCAAGCCCAACCTTGGGACTTGGCGCGGACCGCTTCTTCTTCTTCT	23412
DB	11654	GTGCGCGCGCGCGCGCGCG---TGTGGGGGTGGGGCGGCTTGGGCGCGCTTGGAA	11710
QY	23413	ACCAACCGCGGGAATCATCGACCTTCCACCAACCCCGCGCGCGCGCGCGCGCGCG	23472
DB	11711	ACCGGTGGGGGCTTCTGCTGACCTTGGCGCTTGGCGCGGCTTGGCGCGGCTTGGG	11770
QY	23473	CTCACCCCAACCCCTCACCAACCCCAACCAACCAACCAACCAACCAACCAACCAAC	23532
DB	11771	GCGGTGGAAACGGCTCGCGGTTCGCGGACCAAGATCGCGCTGCGCGCTTGGGAGT	11830
QY	23533	ACCAACACCGCGCGCTTCAACCCCAACCTTACCGCGCGCGCGCGCGCGCGCGCGCG	23592
DB	11831	GCGCGCGCGCTTCAACGAGTTCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGTAC	11890
QY	23593	ACCCCGCGGAAACCACTTCACTACCGCGGACCGCGCGCTTGGCGCGCGCGCGCGCG	23652
DB	11891	CGCGCGCGGACCGGTGCTTGTACCGGTGGACCGCGCGCTTGGCGCGCGCGCGCGCG	11950
QY	23653	CACCACTTCAACCAACCAACCAACCAACCTTCTTCTTCAACCGCGGAAACCGCGCG	23712
DB	11951	CGCTGCTGCGCG---GGCGCGCGCGGAAACCACTTGGCGCTTCAACCGCGCGCGCG	12007
QY	23713	CACACCCCGCGGACCAACCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	23772
DB	12008	GACGCGCGCGCGCGCGCGGACTCGAGGCGGAACTTCTTCTTCTTGGCGCGCGGAGG	12067
QY	23773	ATCACCTTGGGACCGGAAACCAACCAACCTTCAACCTTCAACCAACCTTCAACCTT	23832
DB	12068	TTTGGCGCGTGGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	12127
QY	23833	CCACCAACACCCCTTCAACCAACCTTCAACCGGCGGCTCAATCTTCTTCTTCTTCTT	23892

Db	7964	CTGCGCTCTTGGACGCGCGGTGCGCCAGGAGCGAAGCCCTCTCTGCTACCCGCGGCGTC	8023
Qy	19564	TGGAAGAAATTGGACCGGTCTCTCCAGCAAGTCGTCGCTTCTGCTCGAGGACCTTCCC	19623
Db	8024	CGTCCACCGCCACCGGGACAGGACGAGCGCTCTCGCTGTGGCGCGCTCGTC	8083
Qy	19624	CAGSCACAGGGAATGAGGAGGCGCGAGTACCGTTGAGCAGACGAGAGACAAACCTC	19683
Db	8084	CGGCGCGCGCGCGCGCTGCGCGCGGACGCGCGCGCGAGCGGCGGACAGACGCGGCGC	8143
Qy	19684	CGGCAATC---CTCATGGTTCGGTACGTTCCGAGCAGAGAAAGAGCTGCTCAGCTC	19740
Db	8144	TGGCTGAGCGGGCTCGCGCACAGTCCCGAGGAGCGGCGAGCAGCGCTCAGCTC	8203
Qy	19741	GTCGCAATCCTCCGCGGAGTCTCGGCGGAGCGACTCCGAGGCGCATCCCGCGCGGT	19800
Db	8204	GTACGGGTGTGTCGGGACGCTCTCGGCGACCGCGACTCCCGCGGTTCGGGCGGAG	8263
Qy	19801	CGGCTGTTCAGGATCTAGGTTTCGACTCGCTTTCGGCGGTGAGTTCGCAACCACTC	19860
Db	8264	CGGTCTTCAAGACCTCGGCTTCGACTCTCTGCGCGGCTGAGCTCCGCAACCGCTG	8323
Qy	19861	GCAGCAGACGAGCTGGCTCTGCCGACGACTCTCTGTTTCGATTACCCAGCCCACTC	19920
Db	8324	AACGCGGCCACCGGCTTGGGCTCCCGCGACACGCTCTTCGACCATCCCTCGCGGCG	8383
Qy	19921	AAGCTCGCCCAATTTCTGCTCTCGGAGTCGCGAGTTCAGCCCGGAGTCAACTCCG	19980
Db	8384	CGCTCGCTCCCATCTCTCGCCAGTCCCGGTTGAGAGGGGAGCGGCGGAC	8443
Qy	19981	CTTCC-----CGACCCCGGCGAGAGTTCGATGAGCG	20013
Db	8444	CGACCGTCTGTCGCGGCGGCGCTTCTTCTGCGTACCGTCCGACCGAGCATCCG	8503
Qy	20014	ATCGCATCTTGGCATGGCTGTGCTTCCCGCGGAGTACCTTCGCGGACGACTTC	20073
Db	8504	ATCGCATCTGCGGATGGCATCGCTATCCCGGTGCTGTCTGCGCGGAGGACCTG	8563
Qy	20074	TGGGATCTGATCTCTCCGAGGAGCGGATCGCGGATTCGCCACCGACCGCGCTG	20133
Db	8564	TGGGCTGTGTCGCGGAGGAGCGGACGCGATCAGCGAGTTCCCGCTCAACCGCGCTG	8623
Qy	20134	GACTGAGACGCTCTACGACCCGACCCCGACACCCCGGACCTGCTACACCCGAAAC	20193
Db	8624	GACTGAGAGCTCTACGACCCGATCCCGATCGAAGGACACAGTACTGCCGGAG	8683
Qy	20194	GGCGATTCTCTACGAGGAGCCACTTCGAGCCGAAATCTTCGGCATACGCCCGC	20253
Db	8684	GGCGGTTCTGGAAGGCGCGGTGACTTCGACCGCGCTTCTTCGECATCTCGCGCGC	8743
Qy	20254	GAGCCCTCGCCATGACCCCGACGACCTCTCTCGAAGCCGCTGGGAAACCATC	20313
Db	8744	GAGCCCTGTGATGACCCGAGCGGCTGCTGCTGAGAGTGTCTTCGGAGGCGTG	8803
Qy	20314	GAAACGCGGATCAACCCCGACACCTCTCCAGGACCCCGACCGAGTCTTCACCGC	20373
Db	8804	GAAACGCGGATCGAACCGCTCTCTGCGGCGGAGCGCGGTGTCTACGTGGGC	8863
Qy	20374	ACCAACGAGACAGACCAACCGGACACATCCGTCAGGCCCGGAGCGGTACCGAGGATC	20433
Db	8864	GCCGCGCAACGGCTCTGATCGCTCCGATCCCGGCTGGTCCCGAGGGCTCGAGGGCTAT	8923
Qy	20434	GTCTGACCGGGGACGACAGCATCGCTCCGCGGATCTCTACATCTCGGTTG	20493
Db	8924	CTGCTGACCGGACCGCGGAGGTTGTCGCTCGTTCGCTCGCTCGCTCGCTCTC	8983
Qy	20494	GAGGCGCTCGGCTCACCTCGACACAGGCTGTCTCTCTCGTCTCGCTCGCTCGCTC	20553
Db	8984	GAGGACCGTCCATGAGCGGTGAGAGCGGCTGCTCTCTCTCTCTCTCTCTCTCT	9043
Qy	20554	GCCTGCTGCTCTGAGTCCGGTGAATGCAATGCTGCTTGGCGGCGGCGGCGCTC	20613
Db	10104	AC-----GGCAGGATGCGCTTCTCTTACCGGCGGCGGCGGCGGCGGCGGCGG	10150
Db	9044	GCGGTACGGCGCGCTGCGGACAGCGAGTCTCGCTGCGGCGGCGGCGGCTGCGGTC	9103
Qy	20614	ATGACACCCGATCACTTTCACCGAATTCGCCCGCAACGGGACTCGCCCCCGACGGG	20673
Db	9104	ATGCGCGATCCGCGCGGTTCGTGGAGTTCTCCGCGCAAGAGGGGTGCGCGCGGACGGC	9163
Qy	20674	CGTTCAAGCGGTTCTCGCGCGGCTGACGCTACCGGCTGGGGTGGGGTGTGGGGATG	20733
Db	9164	CGTCAAGCGGTTCTCGCGCGGCTGCGCGCGGACCGGCTGGGCGAGGGCGTCTGCGGTC	9223
Qy	20734	CTGCTGGTGGAGGGCTCTCCGACCGCCCGCGGACCGGCTGCGGCGAGGGCGTCTGCGGTC	20793
Db	9224	CTGCTCTGGAGCGGCTGTCGACCGCGCGCGCGGCGGCGACACGCTCTCTCGGCTCTG	9283
Qy	20794	CGTGGCAGTCCGCTCAACAGGAGCGTTCGAGCAACGGTCTGACCGCGGCGGCGGCGG	20853
Db	9284	ACGCGACCGCGGTTCAACAGGACGCTCTCAACGGGCTGACCGCGGCGGCGGCGGCGG	9343
Qy	20854	TCCAGACGCGGTTCACTCGCGCGGCGCTTCGCAACCGCGGACCTGACCCCGCGGACGTC	20913
Db	9344	GCCAGCAACGCTCATCGCCGAGGCTTCGCGGCGGCTGCGCGGCGGCTGCTCCCGGAGGACGTG	9403
Qy	20914	GATGCGTGGAGGCCCAACCGGACCGGACCACTTTGGGCGACCCGATCGAGGCGGCGGCGG	20973
Db	9404	GACGCGGTTCGAGCGGCAACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9463
Qy	20974	ATCTCGCGACCTACGAGCAGGACGCTCCCGGCAACGGGCGGCTGTTGGGCTGGGCTCGCT	21033
Db	9464	CTGCTCGCGCTTCGCGGACGAACTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9523
Qy	21034	AAGTCCAACTGCGGACACACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21093
Db	9524	AAGTCCAACTGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9583
Qy	21094	ATGCGCTTCGCGGACCGGACACTCCACCGACTCTCCAGCGGAGTACCGCTGCGCGCAT	21153
Db	9584	CAGGCGCTGCGGCGGCGGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9643
Qy	21154	GTGAGTCTGCTCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21213
Db	9644	GCGGCTGAGTCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9700
Qy	21214	GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21273
Db	9701	ACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9760
Qy	21274	GTCTCTCGAAGAGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21333
Db	9761	GTCTCTCGAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9811
Qy	21334	GCGGCTGAGTGGGAGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21393
Db	9812	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9871
Qy	21394	GCAAGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21453
Db	9872	GGAACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9931
Qy	21454	CACCGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21513
Db	9932	ACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	9991
Qy	21514	GACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21573
Db	9992	GACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	10051
Qy	21574	CTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21633
Db	10052	TTGGCGGAGG-----AGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	10103
Qy	21634	ACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	21693
Db	10104	AC-----GGCAGGATGCGCTTCTCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGG	10150

Db 3677 CTCGCTCGCGCGGAGGACCGTATGTCGACACCGGACATGTTCTCGAGTTCTCG 3736
QY 15856 CGGACGCGGCGCTGGCGGTGAGCGGCGGTGCAAGCGGTTCTCGGACGCGGTACGCG 15915
Db 3737 CGGACGCGGCGCTCGCGCGGAGCGGCGGTCAAGCGGTTCTCGGACGCGGTACGCG 3796
QY 15916 ACCGGCTGGGGTGGAGGTGTGGAATGCTGCTGGTGGAGCGGTGTGCGACGCGCGCG 15975
Db 3797 ACCTCTGGGCGAGGCGGTGCGGCTCTCTGTCGTGAGCGGTCTCGGACGCGGAGCG 3856
QY 15976 CTCGGTACCGAATCTCGGGGTGGTGGGAGGTGCGGTCAATCAGACGCGTGGAGC 16035
Db 3857 AACGGCACCGCTGCTGCGGTGATCCGGGCGAGCGGTCAACAGGACGCGGCTCC 3916
QY 16036 AAGGGCTAGCGGCGCCCAACGGGCGTCCAGAGCGGTGATCCGCTGCGGCTGGCG 16095
Db 3917 AAGGGCTACCGGCGCCCAACGGGCGTCCAGAGCGGTGATCCGAGCGGCTGGCG 3976
QY 16096 AAGCGGACCTGACCGCGCGGACGTCGATGCGGTGGAGCGCCACCGCACCGCACT 16155
Db 3977 GACGCGGCTACCGCGCGGCGAGCTGCGCGGTGAGGCGCACGCTACGGGTACCGG 4036
QY 16156 TTGGGACCGGATCGAGGCGGAGCGGCTCTCGGCGACCTAGGACAGGACCGCGCGG 16215
Db 4037 CTCGGGACCGCATCGAGGCGGAGCGGATCTCGGACCTACGCGCGGACCGGGGCGAG 4096
QY 16216 AAGRAACGCTGGGCTGGGCTGATGAGTGAATCGGACATCGGCGGCGGAGCTCGCA 16275
Db 4097 GGGCTCGGCTCAGCTCGGCTCGCTGAAAGTGAACATCGGCGGCGGAGCGCGCG 4156
QY 16276 GGTGGGGGGGTATCAAGATGGTGTGCGGCTCGGAAATGTTCTGTCGCGGAGC 16335
Db 4157 GCGTGGGCGGCTCATAGATGGTCTCGGATGCGGCGCACGCGCTCTGCGGAGCG 4216
QY 16336 TTGATGTGATGAGCGGTGCGGATGAGTGTGCTGCGGGGGCGGTGAGTGTGTG 16395
Db 4217 CTCACGTGGACCGGCGCCACCGCGTCACTGGGAGCGCGCGGCTCGAGTCTCTC 4276
QY 16396 ACGGACGTCCTCGCGCGGCGGAGGCGGCTCGGCGGCGGAGGTGTCATCG 16455
Db 4277 ACGGAGCGGAGTGGCGGA---GACGGGCGCGCGCGCGCGGCGGATCTCTCC 4333
QY 16456 TTCGGGTGAGGGACCAACCGCCACGTCATCTCGAAGAGCACCGCCCAACATC 16515
Db 4334 TTCGGATCAGCGGACCAACCGCCACATCGTGGTGAACAGCGCCCGGAACCGGGAG 4393
QY 16516 CCGTACAGACACCGCGGACGCGCCCGGGAGAGCGCGCGGATGTCGCGGG 16575
Db 4394 GCGGCGGTACCAACCAACCGCGCGGAGGAGGAG---CCGGG 4435
QY 16576 GAAGCGGCGGAGACGCGCGGTACCGGCGGAGGAGGACTGTTCTGCTCGCGGAGT 16635
Db 4436 GAAGCGGCGGAGACCAACCGCGCACCGAGCGCGGCGCGGTCGG-----CGTCCCGAA 4489
QY 16636 CAGGGGTGTGGCGGTGGTGTGTCGCGCAAGTTCGACGCGCGGCGCTGCGCGCCAGGCG 16695
Db 4490 CCGGTACGCGCGCGGTGCTGTCGCGCGGAGCGCGCGCGCTGCGCGCCAGGCG 4549
QY 16696 CAGGCGCTGACCGCGCATCTACCGACACCGCGGCTCGACTCTCGCGAGCTCGGGTAC 16755
Db 4550 GTTCGGCTCGGACCTTCTCGACGCGCGACCGGAGTCACTCGCGGACCTCGGAGCG 4609
QY 16756 ACCCTGCCCGCGCGCGGTGTCGACCGCGGCGGCGGCGGCTATCGCGCGGAGCGG 16815
Db 4610 TCGTGGC---CGCGGTACCGGCTTCGAGCAAGAGCGCGGCTCACCACCGCGGAGG 4666
QY 16816 GACACCTTCTGCAAGCACTTCAGGCACTCGCGGCGGAGGAAACCGCGCGGCTCATC 16875
Db 4667 GACGAGCTGCTCGCGGCTCGACGCGCTCGCGCGGAGGAGCAAGCAAGCGGCTGTC 4726
QY 16876 CACAGAGCGCGCGGAGCGGAGCGGCGGAGCGGCGGAGGCGGAGGAAAGCGCATTCATC 16935

Db 4727 -----ACCGCGCAACCGCGCGGCGGCGGAGCGACGCGGCTTCCTG 4765
QY 16936 TGCTCGGAGAGGACCAACCGCGCGCGCATGGCGCTTACACACCGGCGGCGGCGGCGG 16995
Db 4766 TTACCGGCGAGGAGCGGCGGTGCGCATGGGAGGAACTGCGCGCGGCGGCGGCGGCGG 4825
QY 16996 GTCTTCGCGCGCGCATCAAGACATGTGACCCACTCTCGACCCCGCGGCGGCGGCGG 17055
Db 4826 GTGTTGCGCGCGCGCTGACACCGTGTAGCGGCGCTCGACCGTCACTCGACCGGCGG 4885
QY 17056 CTCCTCCCGCTCTCACCAGGACCGGCAACCGGAGGACCAACCGCTCGAAGAGG 17115
Db 4886 CTCGCGGAGATCGTGGC-----GCC 4906
QY 17116 GCGCACTGCTCAGCAGACCGGCTAGCGCGCGGCGCTCTCGCGCTTCAGGTCGCG 17175
Db 4907 GGGGAGGAGTGGACCTACCGGCTACACCGGCGGCGGCGGCGGCGGCGGCGGCGG 4966
QY 17176 CTCACCGGCTCTCACCAGCGGCTACCACTACCGCGGCGGCGGCGGCGGCGGCGG 17235
Db 4967 CTGTTCCGCTCTCT---CGAACACACCGGCGCTCGTCCCGGACCTGCTCACCGGCGG 5023
QY 17236 CTCGCGGAAATACCGCGCGCGGCGGCTCTCGCGGCGGCTCTCACCGTACCGACCGGCGG 17295
Db 5024 GTCGGCGAGATCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5083
QY 17296 CTCATCACCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17352
Db 5084 CTCGTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5143
QY 17353 CTCACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17412
Db 5144 GTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5203
QY 17413 ATCGCGCGCATCAACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17472
Db 5204 GTCGCGGCGGTCAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5263
QY 17473 CACATCACCGCTCTGCGCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17532
Db 5264 GAGATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5323
QY 17533 GCTTTCCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17592
Db 5324 GCTTTCCTCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5383
QY 17593 CTCACCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17640
Db 5384 CTGACTTTCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5443
QY 17641 CCGGACCACTCTCACCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17700
Db 5444 TCGGCGGAGTCACTCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5503
QY 17701 GCCACCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17760
Db 5504 CTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5563
QY 17761 GACAAACCGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17816
Db 5564 GACGCGGCTCTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5623
QY 17817 -----CTTCACTCTCACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17868
Db 5624 GCGGCGGAGTCACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5683
QY 17869 CTCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17928
Db 5684 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5743
QY 17929 CACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17988
Db 5744 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5803

QY	13696	GCGGTATCATCAGCAGCGCCCGCAGCGGAGCCGAGGAGCCGAGGAAAGACC	13755
DB	1586	GCGGTGTGTACCGGCAACCC-----GCCCGCGGCGCCTC	1621
QY	13756	GCATTATCTGTCTCGGACAGGGGCAACCAAGCGCCCGGAGATGGCCACAGCCCTCTTACCAAC	13815
DB	1622	GCGGTCTGTGTACGCGCCAGGATGGCCCAACGATACGCGGATGGGCAATGGATGTGTACGCC	1681
QY	13816	ACCAACCGGCTCTTCGCGCGCGGACATCAACAGACATCTGACACCACTCGACCCCAACCTC	13875
DB	1682	GCCACACCGGCTTCGGGACGGCTTGGACCGCTTGGACCGATCGCCCGAATCGAACCCCTCTCTC	1741
QY	13876	GACCACCGCTCTCCCTCTCTCACCCAGGACCCCAACACCCAGGACACCAACCCCTC	13935
DB	1742	GACCGGCGCT-----GCGCGAATCT	1762
QY	13936	GAAGAAGCGCGGACTGTCTCAGCAGACCGGTACGCCAGCGCGCGCTCTTCGGCTTC	13995
DB	1763	GTGCGCGCGGCGACACCTCTGACCGGACCGTCCACACACAGCCCGCGCTCTTCGCGCGT	1822
QY	13996	CAGGTGCGCTCTCACCGCTCTCACCGAGGTACACATACACCCCGCACTACTACGCC	14055
DB	1823	GAGGTGCGCTCTCACCGCTCTCT-----CGATCTTGGGGGTCTACGCCCGCACTGTCTGCC	1879
QY	14056	GSACACTCTCTCGCGGAAATCACCGCGCCACCTCGCGGCACTCTCTCACTCTCACCGAC	14115
DB	1880	GGCCACTCTGTGCGGAGATCAGCGCGCCACCGTCTCGCGGGTCTGTCTGTCTGCGCGGAC	1939
QY	14116	GCACACCTCTATACCCAAAGCGGACCTCTCATGTGAACATATGCC-----CCGGGACCC	14172
DB	1940	GCGCGCGCTCTGTGCGCGCGCGCGCTCTATGCAAGCGCTCTCGCGAGGCGCGCG	1999
QY	14173	ATGACACCTCTCACACACCCCAACCAATCACCAACCACTCTCATGACGCGCACCCCGCAAC	14232
DB	2000	ATGCTGCGGTCTGAGGAGCGAGGAGAGTGTCTTCGCACTCTCGGAGCGAGCGG	2059
QY	14233	GACCTCGCCATCGCGCATCAACACCCCACTCTCTCTCATGACGCGCACCCCGCAC	14292
DB	2060	GAGCTCTCTCTCGCGCGGTGAACGGGCGCGCGCGGTCTCTCTCGCGGCGCGGAGCGC	2119
QY	14293	ACGCTCMAACATACCACTCTGTCGCAACAAAGAGCATCAAAACCAAAACCTCTCCCG	14352
DB	2120	GCGGTCTCGACGTGCGGAGCTGCTGCGGAAACAGGGCGCGCGGACGAAGCGGCTCAGC	2179
QY	14353	ACCAACACGCTTCCACTCTCCCGCACCAACCCCACTCTCTCAACCACTCTCAACCAACGAC	14412
DB	2180	GTCTCGCACGCTTCCACTCGCGCTCTATGAGCGGATGTCTGACGACTTTCGCGCGGCTC	2239
QY	14413	ACCAAAACCTCTCACTTACCAACCAACCCCACTCTCTCATACCGGCAACAC-----	14465
DB	2240	GTGAGAGCTTGGATCTTCAGGAGGCGCGGTCTGAGGTCTGTCTCAAGGTGAGGGCGCTG	2299
QY	14466	-----CCACCCGACCAACTCTCTCACCCCGCACTTCTGAGACCAAGAGCCCGCAACCC	14520
DB	2300	CCTGTACAGCGGCGCAATGACACGATCCCGAGTACTGGGTGACCAAGTCTCGCAGCGCC	2359
QY	14521	GTGCACTACGCCACCAACCAACCCCTTCAACCAACAGCGGTCTACCACTTACATCGAA	14580
DB	2360	GTACGCTTCTGACGCGCTTACGACCTCTGAGGAATTCGGGCGCGGACACTTCTCTGGAG	2419
QY	14581	CTCGGACCGGACCAACACCTTCAACACCTTCAACCAACAGCGGTCTACCACTTACATCGAA	14640
DB	2420	CTCGGTCTCCGAGCGGGTCTGTCTCGGATGCGGCGGACTCTGTACGACACAGGAGGCC	2479
QY	14641	ACCACTCTCACTTACCCACCCCGCACCAACCCCGCAACCCCACTCTCTCAACCAACTCTC	14700
DB	2480	GCCACGCGGTCTCCGCGCTTGGCAAGGGCGCGCGGAGCCCCAGTCTGTCTCGCGGCA	2539
QY	14701	GCCAAACCAACCACTTGGACCCCGCACCACTTACCCCAACCAACCAACCAACCCCGAC	14760
DB	2540	CTCACCAACCGTCTTCTGTCTCGGGGCGACGACTGCACTTGGAGCCCGCGCGGAGGACCC	2599

Db 12698 GCTGCGGCGGACC---TCACGCGCCTCGCGCCGCGGAACTCCGACACCGCGCTGCGCGAG 12754
Qy 24490 CTGATTCGACAGGCAATCTGACCGTTCTCGGCTTCGTAATCGGAAGCATCGAGGAC 24549
Db 12755 CTGGTGGAGCGCGGACCGCGGCGCTCGGCTCGACGACCGCGGAGTGGCGGAG 12814
Qy 24550 CAACGAGCCTTCGCGGACCTCGGCTTCGACTCGCTGACGTCGCTCGGCTCAGTTTCAAGAA 24609
Db 12815 GCGGAACGGTTCCCGCGCATGGGCTTCGACTCCCTGGCCACCGTACGGCTGCGCGCGGA 12874
Qy 24610 CTGCGCAAGAAACCGGACCTGACCTCCCGCGTCCCTGCTTCGACTATCCACCCCG 24669
Db 12875 CTGCGCTCGGCCACGGGCTCGACCTGCGCCCGGATCTGCTTCGACCGGGACACCCCG 12934
Qy 24670 CAGGAATGCTGCGCCACTGCGCACAACTCTGTGACCTAGACGACGAAGAGGACGG 24729
Db 12935 GCGCGCTGCGCGCCACTGGCGGAATCTGCTGCCACCGCAGGACCGACCGCGG 12994
Qy 24730 GCACTGTCGAATGCTCTCCCGC 24751
Db 12995 GCGCCCGGACCGTGC CGCGC 13016

RESULT 13

US-09-836-821-1

; Sequence 1, Application US/09836821

; Publication No. US2003087405A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600,438US1

; CURRENT APPLICATION NUMBER: US/09/836,821

; PRIOR FILING DATE: 2001-04-17

; PRIOR FILING DATE: 1998-06-26

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 15872

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-836-821-1

Query Match 7.1%; Score 2188.4; DB 9; Length 15872;
Best Local Similarity 52.5%; Pred. No. 0;
Matches 6973; Conservative 0; Mismatches 5061; Indels 1248; Gaps 50;
Qy 12256 CGCAGCGGAGGTTTCTCTATGACGGCGGCGACTTCGACGCGCGTTCGAGTTGTCA 12315
Db 197 GCGCGGGGGGTTCTCTGACTCCCTCGACTCTTCGACGCGGCTTCGGAATCTCG 256
Qy 12316 CCGCGTGAGCGCTGGCAATGACCGCGAGCGCTGCTGTCGAAACCACTTGGGAA 12375
Db 257 CCGCGTGAGCGCGCGCATGACCGCGAGCGACTGCTCTGAACTCGGCTGGGAG 316
Qy 12376 AGTTTGAACAGGGCGGATCGAACCGGAGTTCATCGCGGGAAGCGGACCGGGGTTTC 12435
Db 317 GCGCTGGAGACCGCGGAATCTGCTCCCGGACCTCGCGGGAAGCGGACCGCGCTCTTC 376
Qy 12436 GTGGGATCAATCCGAGAGCTACACCGGATACACATCATCAGCCCTCAACAGCATC 12495
Db 377 GTGGGACCTCGGAGACCTACAGAGCTCTCTACAGACGGGAGCGGCGCATC 436
Qy 12496 GAGGGTACTGCTCACTGGCAGCGGCAAGCATTCGTCAGCGCGGTATCTCTCAAC 12555
Db 437 ACCAGCACACCATGCGCGGCTGAACCGGGGCTCATCGCCAAACCGCGCTCTGTACAC 496
Qy 12556 TTGGGCTGAAGGCCCTCGGATCACTATCGACACCGCGTGTCTCTCGCTCGTGGC 12615

Db 8564 TGGCGGTGGTGGCGAGGAGAGCGAGTCAGCGAGTTCCCGGTCAACCGCGGTGG 8623
Qy 20134 GACCTGACACGCTCTACGACCCGACCCCGACACCCCGGCACTCTACACCCGAAAC 20193
Db 8624 GACCTGAGAGCCTCTACGACCCGATCCCGAGTCGAGAGGACACAGTACTCGCGGAG 8683
Qy 20194 GCGGATTCCTCTACGACGCGAGGCACTTCGAGCGCCGAATTCCTTCGSCATCAGCCCCCGC 20253
Db 8684 GCGGGTTCCTGGAAGCGCGGTGACTTCGACGCGCCCTTCCTTCGSCATCTCGCGCGC 8743
Qy 20254 GAGCCCTCGCATGGAACCCCGAGCAAGACTCTCTCTCGAAACCGCTCGGAAACCATC 20313
Db 8744 GAGCCCTGTGTGATGGAACCCCGAGCAGCGGTCTGTCTGAGGTGTCTCTCGAGGCGCTG 8803
Qy 20314 GAAACGCGCGCATCAACCCCGACACCTCTCAGCGACCCCGACCCCGAGTCTTCACGCGC 20373
Db 8804 GAAACGCGCGCATCAACCCCGACACCTCTCAGCGACCCCGACCCCGAGTCTTCACGCGC 8863
Qy 20374 ACCAAACGAGACGAGGACGAGGACACATCCGTTCAGGCCCCGAGCGGTACCGAGGATTC 20433
Db 8864 GCGCGACGCGCTCGTAGGCTCCGATCCCGGCTGGTGCCTCGAGGCTCGAGGCTAT 8923
Qy 20434 GTCTGACCGGGGAGGACCAACAGATCGCTCTCGGCGCAATCTCTACATCTCTCGGGTTG 20493
Db 8924 CTGCTGACCGGACGCGGAGCGGCGGTGATGTCCGGCGCATCTCTACGCGCTCGGTCTC 8983
Qy 20494 GAAGGCGCTCGGCTACCTCTGACACAGCGGTTCCTCTCGCTCGCTCGCTCGCTCGCTC 20553
Db 8984 GAAGGACGCTCATGACGCTGAGACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9043
Qy 20554 GCCTGCACTCTCTAGGTCGGTGAATGACCATGCTTCGCGGCTTCGCGCGGCGGCGACGCTC 20613
Db 9044 GCGGTACGCGGCTGCGGACGCGAGTGGGCTCGCGCTGCGGCGGCGGCGGCGGTTG 9103
Qy 20614 ATGACACCCCGATCACTTCAACGAAATTCGCGCGCAACGCGGATCGCGCCCGGACGGG 20673
Db 9104 ATGCGCGATCCGCGGCGGCTTCGTGGAGTTCTCCGCGAGAGGGGCTGGCGCGCGCGC 9163
Qy 20674 CGTTGCAAGCGTTCTCGGCGGCGGTGAGGTACCGGCTGGGTGAGGTGAGGTGAGGTG 20733
Db 9164 CGCTGCAAGCGTTCTCGGCGGCGGCGGACCGGCTGGGCGGCGGCGGCGGCGGCGG 9223
Qy 20734 CTGCTGCTGAGCGGCTCTCCGAGCGCGCGGCAACGCTGCTCTGCGCGGCGGCGG 20793
Db 9224 CTGCTCTGAGCGGCTGTGAGACGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCTG 9283
Qy 20794 CGTGGCAGTGGTCAACGAGACGCTGAGGCAACGCTGCTGACCGCGCGCGCGCGCGC 20853
Db 9284 ACCGCGACGCGGCTCAACGAGGACGCTGCTCAACGCGGCTGACCGCGCGCGCGCGC 9343
Qy 20854 TCCGACGCGGCTCATTCGCGAGGCGCTCGCGCAACGCGGACTGACCGCGCGCGCGC 20913
Db 9344 GCCAGCAACGCTCATTCGCGAGGCGCTCGCGCAACCGCGGCTGTCCTCGGAGGACGCTG 9403
Qy 20914 GATCGGCTGAGGCGCAACGCGGACGCTCTTCGCGGACCGGCTCGAGGCGCGCGCGC 20973
Db 9404 GACGCGCTGAGGCGCAACGCGGACCGGCTCGGCGGCGCGGCGGCGGCGGCGGCGG 9463
Qy 20974 ATCTCGGACCTTACGGAACGACGCTTCGCGGCAACGCGGCGGCTGCTGCGGCTCGG 21033
Db 9464 CTGCTCGCGGCTTCGCGAGGACCGTTTCGCGGACCGGCTGCTGCGGCTCGGCTCGG 9523
Qy 21034 AAGTCAACGCTGGAACACAGGCGCGGCGGCGGCTGCGGAGTGAATCAAGATGGT 21093
Db 9524 AAGTCAACATCGGCTATGCGAGGCGCGGCGGCTGCTGCGGCGGCTCATCAAGATGCTC 9583
Qy 21094 ATGGCCCTCGGACACCGGACACTCCACGAGTCTCCACGCGGATGAGCGCTCGCGCAT 21153
Db 9584 CAGGCGCTCGGACAGGCTTGTGCTGCGCGGCAACCTTCAACGCGGACGAGCGCGCGCAT 9643
Qy 21154 GTGACTGTGCGGCGGCTGAGTGTGAGTGTGAGACGAGTGTGCGGCGGCGGCGG 21213

Db 9644 GCGACTGAGTCTCGCGCGGCTACCGCTCTCACCTCGAGGTGCGTGGC---AGCGG 9700
Qy 21214 GAGGCGCGCGCGCGGCGGAGGTGTCATCTCGGGTTCAGCGGACCAACGCGCCAC 21273
Db 9701 ACCGCGCGCGCGCGGCGGAGCGGGGTTCGCGCTTCGGGTTCGGGCGGACCAATGCCCAT 9760
Qy 21274 GTCATCTCTGAAGAACACCCCGCGAGAGTTCCTCGGGGGGACCAACCGCGGAGGAT 21333
Db 9761 GTGCTCTCGAAGAGCACCCGCGC-----CCGCGCGCGCGGAAACGCGCGGGAG 9811
Qy 21334 GCGGTAAGTGGAGAGAGGTGCTGCGGAGTCTCTGGGGTGGGCGTGGGTGGTGGTGG 21393
Db 9812 GCGCGCGGGGTTCGCGCGCGCAGAAAGGGGCGGAGGCGCCCTGCGCTGGGTGGTCTCC 9871
Qy 21394 GCGAAGTTCGAGCGCGGCTGCGCGCGCAGGCGCGGCGCTGCAACCGCGGACCTCACCGAC 21453
Db 9872 GAGCGGAGAGCGCGGCGCTGCGGTCCAGGCGCGGCGCTCGCGACCACTCTCCCGC 9931
Qy 21454 CACCGCGGCTCGACCTCGCGCGAGTTCGATACACCTCTCGCGCGCGCGCGCGGTTC 21513
Db 9932 ACCCGCGGCGCGCGCGGTGACATCGCTTCTCTCTCGCGCGCGCGCGGAGCCTTT 9991
Qy 21514 GACACCGCGCGACCTCTATTCGCGCGCGGACCGGCAACCTTCTGCAAGACTCTCAGGCA 21573
Db 9992 GACCAACCGCGCGGTGCTGATCGGCTCGGACGCGGCGCGAACTCGCGCGCGCTCGGACGCG 10051
Qy 21574 CTGCGCGGAGGCGAAACCCCGCGCGCTATCCACAGAGCGCGCGCGCGGAGCGG 21633
Db 10052 TTGCGCGAAGG-----ACGCGAGCTCGGCGGTGCGGAGTTCGCGGACCGGG 10103
Qy 21634 ACCGCGGAGCGCGCGAGGAGACCGCATTCATCTGCTCGCGGACAGGCGGACCCAAACCGCC 21693
Db 10104 AC-----GGCAGGATGCGCTTCTCTTACCGGCGGAGGCGAGCGCGCGC 10150
Qy 21694 GCGATGCGCGACCGGCTCTTACCAACCGCGCTTCGCGCGCGGCTCTCAGGCGGCACTCAACGACATC 21753
Db 10151 GCGATGCGCGACGACCTGATCGCGCGGCTCTCTTCTCGCGCTCGCGCTCGGAGGCTG 10210
Qy 21754 TGCACCGACTCGACCG 21813
Db 10211 ACGGACCGCTCTGACCGCGCTGCTCGGCGGCGCGCTCGGCGCGCTGCT-----G 10258
Qy 21814 AACGACAGACGACGAGGAGCGCGCGCGCTTCCAGCAGACCGCGTACGCCCGCGCGCGCGC 21873
Db 10259 GAGCGCGGACCGCGCTCGCGCGCGGAGGCGACTCTTGGACCGGACGAGTACACCGAGCGG 10318
Qy 21874 GCGCTTTCGCTTCGAGTGGCGCTTCCACCGCGCTCTTCCAGCGGCTTACCACTACAC 21933
Db 10319 GCGCTTTCGCGCTCGAGTGGCGCTTCCACCGCGCTGCTG---GAGCACTGGGGGATGCGC 10375
Qy 21934 CCCCACTACTACGCGGACACTCCCTCGGCGGAAATACCGCGCGCGCGCGCGCGCGCGC 21993
Db 10376 CCGGACCTGCTGCTGGGCACTCGGTGGGCGAACTGCGGCGCGCGCGCGCGCGCGCGG 10435
Qy 21994 CTCACCTCACCGAGCG 22053
Db 10436 CTCGATCTCGAGACGCTGCGGCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10495
Qy 22054 CCG 22110
Db 10496 CCG 10555
Qy 22111 ACCG 22170
Db 10556 GCGCGCGCGGAGACCGCGCTGCGCTCGCGCGGTAACCGCGCGCGCGCGCGCGCGCGCG 10615
Qy 22171 AGCGGACCG 22230
Db 10616 TCGCGCGGAGGAGCGGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10675
Qy 22231 ACCAAACCTTCCCG 22290
Db 10676 ACCAGCGCGCTCG 10735

14293 QY ACCGTCAACACATCACACCTCTGCGCAACAGGATCAAAACCAAAACCTCTCCC 14352
14294 DB |||||
2120 DB GCGTCTCGAGCTGCGCGAGCTGTCGCGAACAAGGCGCGGAGCGGCTCAGC 2179
14353 QY ACCAACACGCTTCACATCCCGCCACACAAACCCCATCTCAACCAACTCCACAGCAC 14412
14354 DB |||||
2180 DB GTCTGACGCTTCACATCCCGCTCATGGAGCGATGCTCGACGACTTCGCGCGGCTC 2239
14413 QY ACCCAACCTCACCTACACACCCACACACCCCTCATCACCGCAACAC----- 14465
14414 DB |||||
2240 DB GTCGAAGAGTGAATTCAGAGAGCCCGCTCGAGCTGCTCCACGGTGACCGGCTG 2299
14466 QY -----CCCAACCGACCAACTCTCACCCCTCTACTGTGACCAACCAAGCCGCAACAC 14520
14467 DB |||||
2300 DB CTGTGTCAGAGCGGCGCAATGGACCGATCCGAGTACTTGGTGGAACAGTCCGCGGCC 2359
14521 QY GTGAGTACGCCACACACCAACCCCTCCACAAACAGGGGTACACCTACATCGAA 14580
14522 DB |||||
2360 DB GTACGCTTCTCGACCGCTACGACCTCGAGGAATCGGGCGCGACACCTTCTTGGAG 2419
14581 QY CTGGACCGGACCAACCTCACACCTCACCCACCAACCTTCCCGCAACACCCCAAC 14640
14582 DB |||||
2420 DB CTGGTCCGACGGGGTCTGCTCCGAGTGGCGGAGCTCGTACGCGACCAAGAGGCC 2479
14641 QY ACCACCTTCAACCTCACACCCCAACCAACCCCAACCAACCTCTCTCAACCAACCTC 14700
14642 DB |||||
2480 DB GCCACGGGGTCTCCGCGCTCGCAAGGCGCGCGGAGCCCACTGCTGCTCGCGCA 2539
14701 QY GCCAAACACCAACCTGCGACCCCAACCACTACACCAACCAACCAACCAACCAAC 14760
14702 DB |||||
2540 DB CTACACACCTCTCTGTCGCGGGCCACAGCTCGACTGGACCGCGCGACGGAGCAC 2599
14761 QY ACCCAACCACTCGACCTCCCACTACCTCTCCACCAACCAACCACTACTGCTCCA- 14819
14762 DB |||||
2600 DB GGACCGGTAGGGTGGCTGCGGACTACGCTTCCAGCGCAAGCCCACTGGTTCGAC 2659
14820 QY ACCACCGGCAACCGAGCACCGCTCACCGAGGAGGCGGTGAGCAAGCAACCAACCC 14879
14821 DB |||||
2660 DB GCGCGCGCGAAGCGGGCGCGCTCACGGCGGCGGATCGGGCACCGCTGCGGGCAC 2719
14880 QY ATCAACCCCGCTCGGTATGCTCTGTCGGAAGTCTCGCGAGAGCGAGCAAGACT 14939
14881 DB |||||
2720 DB GCGCGCGCGCGGTGTGACTCGGCGAGGCGAGGCGAGGCGGCGGCGGTGCG 2779
14940 QY GTTGGCGCTG-----GTGCGCAACATCGGCG 14967
14941 DB |||||
2780 DB GGTGGCGGTGATCGGCGGCTCGCCACGAGACGAGCGGTGCGCGCACACGTGCGC 2839
14968 QY GCTGTGTCGGCGCATGCACTCCCGAAGTATGCTTCGGAACAGGCGCTTCAAGAGCTG 15027
14969 DB |||||
2840 DB GCGTCTCGAGTACGACGACCGACCGCGCTCGAATCGGCTCACTTCAAGGAGCTG 2899
15028 QY GGTTTTGTATCTCTGCGCGCAATTCAGCTTCGTAATGAGTGTGTCAGCTTGACCTG 15087
15029 DB |||||
2900 DB GGTTCGACTCTCTCATGTCGTCGAGCTCGGAAACGCGCTCGTCGACGACACGAGCTG 2959
15088 QY CCGCTTCGCGCACGCTGATCTTGATTAACCCCACTCCGATGGGCTT----- 15135
15089 DB |||||
2960 DB CGCTGCCACAGCGGACTGCTTCGACCAACCGCGCGCGCTTCGCGCGCCACCTG 3019
15136 QY TGCAGTTCCTCGGGCGGCGATGCTCGAGCGGACACAGGCAACGACCTCGCTGCGCG 15195
15137 DB |||||
3020 DB GCGCACTGCTACCGCGGCGAGGCGGACCGGATCGCGCGCGGATACCGCGCGG 3079
15196 QY CTAATCGCGTCCCGCGGACGAGCGGATCGCATGCTCGGCAATGGCTTTCGTAACCC 15255
15197 DB |||||
3080 DB ACCCGCGGACACCAACCGCGGACCAATCGCGATCATCGGCTTGGCTTCCGCTACCC 3139
15256 QY GGTGATGACGAGCGGTGATATCTTGGCAGGTGTGATGTTGGCCATGACGCGATC 15315
15257 DB |||||
3140 DB GCGGGCGTCACTCCCGGAGGACCTGTGCGCGCTGTGCGCGGCGGCGAGCGGCTC 3199

15316 QY GCGGATTCGACGAAACCGTGGTGGGACCTCGACACGCTGTATCAACCCCGACCCGAGC 15375
15317 DB |||||
3200 DB TCGGGCTGCGCAACCGACCGCGCTGGGACGAGAC---CTCTTCGACCGCGACCCGAC 3256
15376 QY CACACGGAACAGCTACACCCGAGCGGCGGATTCCTTTACGAGCGGAGCAATTCGAT 15435
15377 DB |||||
3257 DB GCGAGCGCAAGAGCTCGGTCCGAGAGGCGGATTCCTGACACGACCGCGCTCTGTCGAC 3316
15436 QY CCCGACTTCTTCGCTATCAGTCCGCTGAGGACACTGGCGATGGACCCCGACGAGCGGCTG 15495
15437 DB |||||
3317 DB GCGGCTTCTTCGGATATCGCCCGGAGGCGCTTCGGCATGGACCCCGACGAGCGGCTG 3376
15496 QY TGTCTGGAACAGCGTGGAGAGCATGAAACACGCTGATCAACCCCGACGAGCTCTCGT 15555
15497 DB |||||
3377 DB CTCTTGAGACGCGCATGGAGGCGCTGGAGCGCGAGGCTCGACCCCGAAGGCTCAAG 3436
15556 QY GGCACACCAACCGGGCTTTCGCGGGCTGACCTACACAGCTACGCGCGGCTTTCCTCC 15615
15557 DB |||||
3437 DB GGCAGCGGACGCGCGCTTCTGTCGCGCCACCGCTTGGACTACGCGCCGCGCATGAC 3496
15616 QY ACAGCTCCGCGAGGGTTCGAGGGTATCTCGGGCACGGAAGCGAGGAGTATCGCTCG 15675
15617 DB |||||
3497 DB GACGGCGGAGGGGTGAGGGCCACTCTCTGACCGGACCAACCGCCAGCGTGTATCG 3556
15676 QY GGTCTGTCGCTTACGCTCTCGGCTGGAAGTTCGCGCTTCAAGTTCGACATCGCTGCG 15735
15677 DB |||||
3557 DB GCGCGCATCGCTTACAGCTCGGCTTACCGCTTCTGCGGTACCGTTCGACACGCGCTG 3616
15736 QY TCTTGTCTCTGCTGCTCTGACCTGCGCTTTCAGGCGCTGCTCGGCGGAGTTCCTCC 15795
15737 DB |||||
3617 DB TCGTCTCTGCTGCTGCGCTGACCTGCGCTTCTGCTGCGGACGCGGAGTTCGAGC 3676
15796 QY ATGGCGCTTCGCGGGTGGGTGATGATGATCAACCCCGCGCGGTTCGTGGAGTTCG 15855
15797 DB |||||
3677 DB CTCGCGCTTCGCGGCGGAGCACCGCTGATGCGACACCGGCAATGTTCTGAGTTCG 3736
15856 QY GCGCAGCGGGCTTCGCGTGGAGCGGCGGTGCAAGGCGTTCGCGGACGCGCTGACGCG 15915
15857 DB |||||
3737 DB GCGCAGCGGGCTTCGCGCGGACGCGCGCTTCAAGGCTTCTTCGACTCTCGCGACGCG 3796
15916 QY ACCGCTGGGTGAGGGTTCGGAATGCTGCTGGTGGAGCGGCTGTCGAGCGCGGCGG 15975
15917 DB |||||
3797 DB ACCTCTGGCGGAGGGGCTGCGCTTCTGCTGTCGAGCGGCTCTCGACCGCGGCGC 3856
15976 QY CTCGCTACCGAATCTCTCGGCTGCTGCTGAGTGGCTGAGTGGCTCAATCAGGACGCTCGAGC 16035
15977 DB |||||
3857 DB AACGCGCAACCGCTGCTGCGCTGATCCGGGCGAGCGGCTCAACAGGACGCGCGCTCC 3916
16036 QY AACGGCTGACGCGCGCCCAACGCGGCGCTCCAGAGCGGTGCTATCCGCTGCGCTTGGCC 16095
16037 DB |||||
3917 DB AACGGCTCACGCGCGCCCAACGCGCGCTCCAGAGCGGCTATCCGACAGGCGCTTGGCC 3976
16096 QY AACGCGGACCTGACCGCGCGCGGCTGATGCGGTGGAGGCGCCACGCGACCGGCACT 16155
16097 DB |||||
3977 DB GACGCGGGCTCACCGCGCGCGGCTGACGCGCTGCGAGCGGCTGCGAGGCGGACGCTG 4036
16156 QY TTGGCGGACCGGATGACGCGCGCGGCTTCTTCGCGCACTACGAGACAGGACCGCGCGC 16215
16157 DB |||||
4037 DB CTCGCGGACCGCATGAGGCGCGGCGGATCTCTCGGCACTTACGCGCGGACCGCGGCGGAG 4096
16216 QY AACGAACCGCTGTGGCTGCGATGAAGTGAACATCGGCGCAACGCGAGGCTGCGCGCA 16275
16217 DB |||||
4097 DB GCGCTCGCTCCAGCTCGGCTCGTGAAGTGAACATCGGCGCAACGCGCGGCGCGCGG 4156
16276 QY GGTGTGGCGGGGTATCAAGATGCTGATGCGCTGCGGATGCTGCTGCGCGGCGGAG 16335
16277 DB |||||
4157 DB GCGGTGGCGGGCTCATCAAGATGCTGCTGCGGATGCGCGCAACGCGGCTCTGCCCCAGGAG 4216
16336 QY TTGCAATGAGTGAAGCGCTGCGCGCATGTGAGTGTGCTCCGCGGCGGCGGCTGAGCTGCTG 16395
16337 DB |||||
4217 DB CTCACGTGAGACCGCGCGCCACACCGCGCTGACTGGAGGCGCGGCGGCTGAGCTCTC 4276
16396 QY ACGGAGACGGTGCCTTCGCGCGGAGGCGGCGGCTGCGCGGCGGAGGAGTGTATCG 16455

; NUMBER OF SEQ ID NOS: 53

; SEQ ID NO 1

; LENGTH: 15872

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-988-384B-1

Query Match 7.1%; Score 2188.4; DB 9; Length 15872;
Best Local Similarity 52.5%; Pred. No. 0;
Matches 6973; Conservative 0; Mismatches 5061; Indels 1248; Gaps 50;

QY 12256 CGAGCGGAGGGTTTCTCTATGACGGCGGGCGACTTCGACGCGCGTTCCTTCGAGTTGTCA 12315
DB 197 CGCGGGGGGGTCTCTCGACTCCCTCGAATCTTCGACGCGGCTTCTTCGGAATCTCG 256
QY 12316 CCGCGTAGCGCTGGCAATGGACCGCGACGCGCTGCTTCGAAACCACTTCGGAA 12375
DB 257 CCCCGTAGCGCGCGCATGGACCGCGACGAGCTGGTCTCGAATCTCGCTGGAG 316
QY 12376 ACCTTCGAAACAGGGCGGAATCGACCGAGGTCATGCGGAAACCGGACCGGGTTTC 12435
DB 317 GCGCTGGAGGACGCGGGAATCGTCCCGGACCTCGCGGAAGCGCAACCGCGTCTTC 376
QY 12436 GTGGGATCAATCCGAGGACTACACACCGGATACACATCAGCCCTCAACGCGATC 12495
DB 377 GTCGGACACCTCGGGACGACTACAGAGCTCTTACCGACGCGCGAGCAGGCGCATC 436
QY 12496 GAGGGTACCTGCTCACTGCGAGGGGCAAGCATTCGCTCAGGCGGTATCTCTACAAC 12555
DB 437 ACCAGACACCATGGCGGGCGTGAACCGGGGCGTCATCGCAACCGCGTCTCGTACCAC 496
QY 12556 TTCCGGCTCGAAGGCCCTCGCATCTATCGACACCGCGTGTTCCTCTCGCTCGTCGCG 12615
DB 497 CTCGGCTCGAGGGCCGAGCTCACCGTCGACCGCGGAGTCTCTCGCTCGTCGCGC 556
QY 12616 CTGCTCTGCGCTGCGAAGCGCTTCGCTCGGTGAATGACCATGCGCTTCGAGCGCG 12675
DB 557 GTCACCTGCGCTGCGAGTCCCTGCGCGCGGGAGTCCACGACGCGCTCGTCGCGCGC 616
QY 12676 GCCTCGCTATGCGCACTCCCTCTGCTTTCACCGAGTCTCTCGCAGCGGGCTCGGC 12735
DB 617 GTGAACCTCAACATCTCGCGAGAGCGCGTGAACGAGAGCGCTTCGCTGAGCTCTCC 676
QY 12736 GCACAGCGCGGTGCAAGCGTTTTCGCGGGCGGCGGACCGGACCGGCTGTCGAGGCT 12795
DB 677 CCGGACGGCACCGCTTACACTTCGACGCGCGGGCCAAACGGATTCTTCGCGGGCGAGGC 736
QY 12796 GTGGGATGCTGTGTGAGCGGCTCTCGACGCGCGCGCAACCGGTACCGTGTCTGT 12855
DB 737 GCGGAGTGTCTGTACTCAAGCGCTCTCCGCGCGCTCGCGACGCGACCGGTGTCCAC 796
QY 12856 GCGCTGTCGGGCGACGCGCTCAACAGGAGCGGCGCAAGCAACGCGCTGACCGCACCC 12915
DB 797 GCGGTATCGCGCCACGCGCGCTCAACAGAGGAGCGCACCGCGGTCTCAACCGTGC 856
QY 12916 AAGCGTGTTCACAGTCAAGGTCAATCCGCGAGGCTTTGGCCAAACGACACCTCTCCCT 12975
DB 857 AGCAGGGCGCCAGAGAGGTGTCTCGGAGGCGTACGAGAGCGGCGCTTGACCCG 916
QY 12976 GCGGATGTGATCGGTGAGGCGCCACGGAACCGGACCACTTCGCGGCGACCGATCGAG 13035
DB 917 TCGCGCTCGATGATCGTGAATCTCAACGCGACCGGAACCGCGCTCGGCGACCGCATCGAG 976
QY 13036 GCTCAAGCCCTGTCGAACTACGTCAGGACCGCGCCCAACGCGCGCGCTCTGGTTC 13095
DB 977 GCGCGCGCTCGCGCGCTCTCTCGCTCGCGCGCGCGCGGACGAAACCTGTCTGCTTC 1036
QY 13096 GGAACCTCAAGTCAACATCGGGCACTCATGGCGCTGCGGCTGTGGCGGGGTCTATC 13155
DB 1037 GGTCTGGCCCAAGACGAACTCGGGCACTTGAAGGCGCGCGCGCATGTCTGGCTCATC 1096
QY 13156 AAGATGGTATGCGCTGCGGAATGGTGTGCTGCGCGGACGTTGATGTGATGAGCGG 13215
DB 1096 AAGATGGTATGCGCTGCGGAATGGTGTGCTGCGCGGACGTTGATGTGATGAGCGG 13215

DB 1097 AAGACGCTCTCTCGCTCGCGCGCGCGCGGATCCCGGCGAGCTCAACTTCGTACGCC 1156
QY 13216 TCGCGCATGTGACTGGTTCGCGGTGCGGTGCGAGTCTGACGAGAGCGTGCCTCG 13275
DB 1157 CACCGGACATCCGCTCGACACCTCGGCTCGAGCTGCCGACGCGCTCGGGAGTGG 1216
QY 13276 CCGCGGGGAGGGCGGCTACGCGGGGAGGAGTGTCTATCTATCTCGGCTCGAGCGGAC 13335
DB 1217 CCGACCCCGGAC--CGCGAACTCTCTCGCGCGGTGAGTCTGTCGATCGGCGGAC 1273
QY 13336 AAGCGGACGCTATCTCTGAGGAAGACCGCGCGCAACATCCGCTCAGACACCGCGC 13395
DB 1274 AAGCGGACGCTCTCTCAGCGAAGCGCGCGCGCG-- 1309
QY 13396 GACGAGTCCCGGAGAAATCAGCCCGACGAGGATGCGGTAGTGGCGATGAGGCTGT 13455
DB 1310 -----GCGCGGAGAGCGCGGATCGATGAGGAGACCC 1345
QY 13456 GCGCGAGTCCAGGGGTGTGGCGGTGTGGCGCAAGTTCGAGCGCGGCGCTTCGCG 13515
DB 1346 GTGACAGCGGGCGCGACTGCTCTCTGTCGTACCGCGCGCGGCGGAGCGCTCGC 1405
QY 13516 GCGGAGCGGCGGCGCTGACGCGCGCACTCACGACACCGCGGCTCGACCTCGCGGAC 13575
DB 1406 GCGGAGCGGCGGCTGACGAGGCGGTTCGAAGCGGACCGCGAGTTCGCGCGCGCGCA 1465
QY 13576 GTGGGTACACCTCGCGCGCGCGCGGTTCGACCGCGGCGCGGAGCGCGGAGAGAC 13635
DB 1466 CTCGCGCGTCTGCTGCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1525
QY 13636 GCGGACCGCGACACCTCTCTGCAAGCACTTCAGGCACTTCGCGCGGCGGAAACCCAC 13695
DB 1526 CCGGACCGCGCGCGCTCTCTGCGCGCTTCGCGCGCTTCGCGCGGCGGCGCGCGCG 1585
QY 13696 GCGGTATCCACAGAGCGCGCGCGGAGCGGAGCGGAGCGCGGAGCGCGGAGAGAC 13755
DB 1586 GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1621
QY 13756 GCATTCATCTGCTCGGACAGGCGACCGACCGCGCGCGGCGGCGGCGGCGGCGGCG 13815
DB 1622 GCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1681
QY 13816 ACCACCGCGCTTCGCGCGCGCACTCAACGACATCTGACCGCACTTCGCGCGGCGGCG 13875
DB 1682 GCGGACCGCGCTTCGCGCGCGCTTCGCGCGCTTCGCGCGGCGGCGGCGGCGGCGG 1741
QY 13876 GACGACCGCGCT 13935
DB 1742 GACCGCGCGCT-----CGCGGAGCTC 1762
QY 13936 GAAGAAGCGCGCGCACTGCTCCAGCAGACCGCTAGCGCGCGCGCGCGCTCTTCGCG 13995
DB 1763 GTCGCGCGGCGGCGACACCTCGACCGCGCGCTTCACACAGCGCGCGCTCTTCGCG 1822
QY 13996 CAGGTGCGCGCTTCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14055
DB 1823 GAGGTGCGCGCTTCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1879
QY 14056 GGAACCT 14115
DB 1880 GCGGACCT 1939
QY 14116 GCGGACCT 14172
DB 1940 GCGCGCGCGCT 1999
QY 14173 ATGACACCT 14232
DB 2000 ATGGTCT 2059
QY 14233 GACCT 14292
DB 2060 GAGTCT 2119

22879 TCCTGCTGACCGCCACCAACACGACCAAGCCCGCATCAACACACTTGGACCTTACCAGGAA 22938
Db CGCGCTACTGCTCGCCCGGCGCGCGGGAACACGACGACTGGATGTACCGATC 11233
22939 ACTTGAACCCCT----CACCCTCCACACACCAACCAACCAACCAACCTTGGCTATC 22995
Db GGTGGGACCGGTGCGCGCTGTGACCGCGCGGCGCGGACCGCGCGCTTGGCTGGTG 11293
22996 GCATCCCCGAAACCCAGACACCAACCCACATCAACCAATCTTCAACCACTTCCAC 23055
Db ATCCACCCCGACAGCCCGGCTGCGGGAGCTGTCCGGCAACCGCGGCTGCGC 11353
23056 CACACGGGATACCCCCATCCCTCTCACTGTCAACCAACCAACCAACCAACCAACCAAC 23115
Db GCGCGGCGGAGCGCCGTACCGCTGCGCGTGGACGCTCCGGCGCGCACCGCGGCTCC 11413
23116 CTCACCAACCTTCAACA----CACCAGCAACAGCCCAACCAACCAACCAACCAACCC 23172
Db TTCCGGGCACTGCTCGGCTCCCGCACCGACTGACACGAGGTGACACAGCCCGGCCC 11473
23173 ATCACCGGCTGCTCTCTCTCGGCTCGACGAAACACCCACCCACCAACCCACCCAC 23232
Db GTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11533
23233 ACACCCACCGGACCTCTCTCAACCTCACCTCCCCCAACCAACCAACCAACCAACCA 23292
Db GTACCGCGCGGCTCTGCGACGCTGCTCCCTGATGAGCTATGAGGAGGCGGTG 11593
23293 CCNACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 23352
Db GAGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11653
23353 ACACACCCACCAACCAACCTGCGGACTCGCCCGCACCAACCTCTCTGAAACACCC 23412
Db GTGGCGGCGGCGCGCC---TGTTGGGGCTGGGGGGGGTGGCGCTTGGAAACCC 11710
23413 ACCACACCGCGGGAATATGACCTTCCCAACCAACCAACCAACCAACCAACCAACCA 23472
Db ACCCGTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11770
23473 CTCACCAACCTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCC 23532
Db GCGTGGAAACGCTCGCGGCTCCGAGGACCAAGTCCGCTGCGCGCTCGGCAAGTTGG 11830
23533 ACCACACCGCGGCTCACCCACCAACCTTCAACCAACCAACCAACCAACCAACCC 23592
Db GCGCGGCTCACAGGCTGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAC 11890
23593 ACCCCACCGGAACCACTTATACCGGGGAAACCGGGCGCTCGCCACCAACCTTACC 23652
Db CGGCGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11950
23653 CACCACTTACCAACCAACCAACCAACCAACCAACCTTCTCTTACAGAGCGGAACCG 23712
Db CGTGGCTCGCCG---GGCGGCGCGCAACCACTTGGCGCTACCAACCGCGCGGCGCG 12007
23713 CACACCCCGCACCAACCTTCAACCAACCTTCAACCAACCAACCAACCAACCTTACC 23772
Db GACGCGCGCGCGCGCGGACTGAGGCGGAACCTTCTCTCTTGGCGCGCAAGTGAGC 12067
23773 ATCACCACTTGGACACCAACCAACCAACCAACCAACCAACCTTCTTCAACCACTTCC 23832
Db TTGCGGCTTGGACACCGCGGACCGGCGGCTTGGCGGCTTGGCGGCGGCTTGGCGG 12127
23833 CCACACACCTTCAACCAACCTTCAACCAACCAACCAACCAACCAACCTTCTTGG 23892
Db GAGGACACCCCGCTTACCGGCTTCTTCAACCGCGGCGGCTTACCGAGGTACCGCGCT 12187
23893 TCGGAAACCGATCGCGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 23952
Db TCCCGTACTTCCCGCGGACCTTCCCGGAGGTGACCGGCGGAGCGGCGGCGGCGG 12247
23953 ATTCTGATGAGTTGCTGCTGGACCATGAAC---GCTTGAACACTTCTTCTTCTTCT 24009

Db CACCTGGACGAATGACCCCGGAATCTCGCGCCCGGACTCGACGCGTTCTCTACTCC 12248
24010 TCGGCGCGCGGCTTGGGCGAGCGGAATCAGTGCATACTCGGCGGCAACGCAATAC 24069
Db TCGGCGCGCGGCTTGGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12367
24070 CTGACGCGCTCGGAGACCATCTGTCAGACACATGGAATCTCCCGGGGATCGATCGCCTGG 24139
Db CTGACGCGCTCGCCCGCGCGCTGCGGCGGAGCGACTCCCGCGCACCTCTCATCGCCTGG 12427
24130 GCGCCCTCGGCGGAAAGGCGATGTGCGCGGCTGATGCGGCTCATGGTTACTTGAAGAAG 24189
Db GCGCTGTGCGGCGGCGGCGGTATGCGGCGCGACGAGGCGGCGGCGGAGTATCTGCGCGG 12487
24190 GCGGCACTTCTCGCATGAGCGACCATGCGCTGCGGCGATTCATCTGCTGCGGCGG 24249
Db CCGGTATGCGGCGCATGCGACCGGTCTCGCGCTTCCGGCGATGGCCACCGCATCGCC 12547
24250 CAGCGCGCAATTCCAACTGATCATCGGACATCGACTGCGGAGCGCTTCTGCTCCCGCC 24309
Db TCCG---GGAACCTGCGCCACCGTCAACCCACCGACTGGAGCGCTTCGGCGAGGCG 12604
24310 TTCACCGCTCGACGCGCACAGCGCGCTCATCGAGGACATTCGGAGGTTTGGGCAAGCGCT 24369
Db TTACCGGCTTTCGGCGCGGCGCTCTGATCGCG-----GCTCGGCGACCGCGG 12654
24370 CAGGAGCTTGGAGGAGCTGCTGCGAGCGCAAGACACAGCTCAGCGGATTCGCGAGC 24429
Db CGCGGCGG-----GGCGGCGGAGACCCCGAGGAGGGAACGCCACC 12697
24430 TCTCTCCGTGAGCGATTGGCGCGACTGACGCTCTCAAGCAGAACAGGTGCTGCTCGGC 24489
Db GCTCGGCGGAGC---TCACCGCGCTGCGCGCGCGGAACTCCGACCGCGCTGCGCGAG 12754
24490 CTGATTGCGACAGGATCTGACCGCTTCTCGGCTTCTGATTCGGAAGGACATCGAGGAC 24549
Db CTGTGCGAGCGCGGACCGCGCGCTGCGCTCGACGACCGCGCGGAGTTCGCGGAG 12814
24550 CAACGAGCTTTCGGGAGCTGCGGCTTCTGACTGCGTACGCTGCGCTCAGTTCAGCAAGAA 24609
Db GCGGAAACGCTTCCCGCGCATGGGCTTCTGCTTCTGCGGCGGCGGCGGCGGCGG 12874
24610 CTCGCAAGGAAACCGGACTGCGCACTCCCGCGCTCCCTGCTTCTGACTATCCCAACCGG 24669
Db CTGCGCTCGGCGACCGGCTCGACCTGCGCGCGCGGATCTGCTTCTGACCGGAGACCGCG 12934
24670 CAGGAATGTGCTGCCATCTGCGGACACAACTGCTGACCTTAGACGAGGAGGAGCGG 24729
Db GCGCGCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12994
24730 GCACCTGCTGAATGCTCTCCCGC 24751
Db GGCCCGGCGGCGGCTGCGCGC 13016

RESULT 12

US-09-988-384B-1

; Sequence 1, Application US/09988384B

; Publication No. US20030073824A1

; GENERAL INFORMATION:

; APPLICANT: Sherman, D.H.

; APPLICANT: Liu, H.

; APPLICANT: Xue, Y.

; APPLICANT: Zhao, L.

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin

; FILE REFERENCE: 600.536US1

; CURRENT APPLICATION NUMBER: US/09/988,384B

; CURRENT FILING DATE: 2001-11-19

; PRIOR APPLICATION NUMBER: PCT/US99/14398

; PRIOR FILING DATE: 1999-06-25

; PRIOR APPLICATION NUMBER: US 09/105,537

; PRIOR FILING DATE: 1998-06-26

Db 9164 CGTGCAGAGCGTTCTCGGCGCGCCGACGCGACCGGCTGTGGCCGAGGCGCTGCGCGTG 9223
Qy 20734 CTGCTGTGAGCGGCTCTCCGACGCGCGCGGAAACGGTCAACGCTGTCTTGGCGCGTG 20793
Db 9224 CTGCTGTGAGCGGCTGTGAGCGCGCGCGCGGCGGACAGGTTCTTGGCTTGGTC 9283
Qy 20794 CGTGGCAGTCCGCTCAACAGGACGCTGTGAGCAACGCGTGTGACCGCGCCCAACGCGCCC 20853
Db 9284 ACCGGCACCGCGCTCAACAGGACGCTGTGCTCAACGCGGCTGACCGCGCCCAACGCGCCA 9343
Qy 20854 TCCGACGAGCGCTCATCCGCGAGGCGCTCGCAACGCGGACCTGACCCCGCGGAGTC 20913
Db 9344 GCCCAGCAACGCGCTCATCGCGAGGCGCTCGCGAAGCGCGGCTGTCCCGAGGACGTC 9403
Qy 20914 GATCGGTGAGGCGCCACGCGACCGGACCACTTTTGGGCGACCCGATCGAGGCGCCAGGCC 20973
Db 9404 GACCGGTGAGGCGGACGCGACCGGACCCGCTCGGCGACCCCATCGAGGCGCGGGCG 9463
Qy 20974 ATCCTCGGACCTACGAGACGAGCCGCTCCGCGCAACGCGGCTGTGTGGCTCGGCTCCGTC 21033
Db 9464 CTGCTCGCGCTCCGAGCGAAACCGTTCCGCGCACCAACCGCTGTGGCTCGGCTCGCTG 9523
Qy 21034 AAGTCCAACTCGGACACACAGGCGCGCGGCGGTGCGCGAGTGATCAAGATGGTG 21093
Db 9524 AAGTCCAAATCGGCGATGCCAGGCGCGCGCGGTGTGCGGCGCGCTCATCAAGATGCTC 9583
Qy 21094 ATGGCCCTCGCACCGGACACTCCGACGACTCTCCAGCGGATGAGCGCTGCGCGCAT 21153
Db 9584 CAGCGCTGCGGACCGCTTGTCTCCCGCACCCCTTCACGCCAGCGGACCCCGCAT 9643
Qy 21154 GTGGAATGCTCCGCGGTGCGGTGCTGCTGACGAGACGCTGCTGCGCGCGCGG 21213
Db 9644 GCCGACTGAGCTCCGCGCGGTACGCTGCTCACTCCGAGGTGCGGTGGC 21213
Qy 21214 GAGGCGCGCGCGGCGGAGAGTGTATCATTTGCGGTGAGCGGACCAAGCGCCAC 21273
Db 9701 ACCGCGCGCGCGCGGAGACCGGCGGTGTCGCGCTTCGCGGTGCGGCGGACCAATGCCAT 9760
Qy 21274 GTCATCTCGAAGAACCCCGCGAGCGATTCGCGGCGGACACCCCGCGAGGAT 21333
Db 9761 GTGCTCTCGAAGGACCCCGC-----CCGCGCGCGCGGAAACCGCGCGGGAG 9811
Qy 21334 GCCGTGAGTGGGAGGAGGTGTGCGCGGAGTCTTGGGAGTGTGGCGTGTGTGTG 21393
Db 9812 GCCCGCGCGCTCCCGCGCGCAGAGGCGGAGGCGCCCTGCGCTGGGTGTCTCC 9871
Qy 21394 GCCAAGTGGAGCGCGCGCTCGCGCGCGAGGCGGCGGCGCTGACGCCCACTCACCGAC 21453
Db 9872 GGACGCGACGAGCGCGCGCTCGCGGTCCAGGCGCGCGCTCGCGACCACTCTCCCGC 9931
Qy 21454 CACCCCGGCTCGACCTCGCGAGCTCGGATACACCTCGCCACGCGCGCGCGGTTC 21513
Db 9932 ACCCGCGGCGCGCGGTGATCGCTTCTCCCTCGCGCGCGCGCGCGAGCTTT 9991
Qy 21514 GACACCGCGCCACCTCATCGCGCGGACCGCGACACCTTCTGTGAAGCACTCCAGGCA 21573
Db 9992 GACCAACCGCGCTGTGATCGCTCGGACGCGGCGCACTCGCGCGCGCTGAGCGG 10051
Qy 21574 CTGCGCGAGGGAACCCCAACCGCGCTCATCCAGAGCGCGCGCGGAGGACCGG 21633
Db 10052 TTGGCGGAAG-----ACGCGAGCTCCGCGGTGTGGCGGAGTCCGCGACCGG 10103
Qy 21634 ACCGGGAGGCGCGAGGAAGACCGATTCTCTCGGACAGGCGGACCCAGCGCC 21693
Db 10104 AC-----GGCAGGATGGCTTCTTCCCGGAGGCGGACCGAGCGCGC 10150
Qy 21694 GGCATGGCCAGCGCTCTACCAACCCACCGCGCTTTCGCGCGCGCACTCAACGACATC 21753
Db 10151 GGGATGGCCCAAGACTGATGCGCGCCATCTTCTTCTGCGCTCGCGCTCGACGAGGTG 10210
Qy 21754 TGCACCACTGACCCCACTGACACCGCTCTCTCCCGCTCTCAACCAAAACGAC 21813
Db 10211 ACGGACCGTCTCGACCCGCTGCTCGGCGCGCGCTCGGCGCGTGT-----G 10258

Qy 21814 AACGACAAACGAAACGAGACGCGCGCTCTCCAGACAGACCCCGTACGCCAGCCC 21873
Db 10259 GAGCCCGAGCCCGGCTCGCCGAGGCGACTCTCTGGACGAGCAGGATACACCGAGCG 10318
Qy 21874 GCGCTTTCGCTTCAGGTGCGCTTCCACGCGCTCTCACGAGCGGTACACACATCACC 21933
Db 10319 GCGCTTTCGCTTCAGGTGCGCTTCCACGCGCTCTG---GAGCACTGGGAGTGGC 10375
Qy 21934 CCCCACTACTACGCGGACACTCCCTCGGGAATACCGCGCGCCACCTCGCGCGCATC 21993
Db 10376 CCGGACCTGCTGCTGGGCACTCGGTGGCGAACTGGCGCGCCACCTCGCGGCTG 10435
Qy 21994 CTCACCTCACGAGCGGACCAACCTCATCAACGCGCGCACTCTCATGAAACCATG 22053
Db 10436 CTGATCTCGACGACGCTGCGGTGTGGCGCGCGCGGAGGTGATGAGCGCTG 10495
Qy 22054 CCCCC---CGGCAACATGACACCTTCCACACACACCAACATACACCAACCATC 22110
Db 10496 CCGCGCGCGCGCTGCTCTCGTGGCGCGGAGGACGAGTCCGCGCACTGCTG 10555
Qy 22111 ACGGCGCAGAAACGACTCGCATCGCGCATCAACACCCCACTCCCTCTGCTCATC 22170
Db 10556 GCGCGCGCGGAGCGCGCTCTGCGTGGCGCGGTGAACGCGCGCGGCTCGGTGCTGATC 10615
Qy 22171 AGCGGCAACCCCGCACACGCTCAACACATCACACCTCTGCAACCAACGAGCATCAA 22230
Db 10616 TCGGCGGAGGAGCGGTGCGCGGCGGCGGAGCTCGCGGACGAGCGCGCGC 10675
Qy 22231 ACCAAACCTCTCCCAACACACGCTTCCATCTCCCGCACCAACACCAACCTCTCAAC 22290
Db 10676 ACCAGCGGCTCGCGTTCGCGCACGCTTCCACTCACCTGATGAGCGCATGCTCGCC 10735
Qy 22291 GAATCTCAGCAGCACACCAACCTCACTACGACCAACCCCAACACCCCTCTCATCAC 22350
Db 10736 GGATTCGCGGAGTCCGCGCGCGCTTACCGGGAACCGGAGTGAAGTGTCTCTCC 10795
Qy 22351 GCCAACA-----CCGACCGGACCAACTCTCACCCCGCACTACTGAGCCCAA 22398
Db 10796 AGGTGACGCGGCGCGCGCGCGCGCGGTAACCTACCGGCGCGGACTACTTGGTGGC 10855
Qy 22399 CAAAGCGGCAACACGCTGACGCTACGCGCACACCAACCAACCTTCCCAACGAGCGCTC 22458
Db 10856 CAGTTCGCTGAGCGCGCTTTCGCGGAGCGGTCGCGACGCGCACCGCTCGGAGC 10915
Qy 22459 ACCACTTACATGAACTCGGACCGCGAACAACCTTACGACCTCACCGACCAACCTCTC 22518
Db 10916 CGCACTTCTTGGAGACCGCGCGCGGCTGTGTGCGGATGGCAGAGGTGCTCTG 10975
Qy 22519 CCCAACACCCCGCACACCTCACTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22578
Db 10976 GAGGAGACACGCTGCGCGCTTGTGCGGCGATCAAGACCGCGCACCGCG----- 11026
Qy 22579 CTCCTTACCACCTCGCCAAACACACACCTGCGGCGCGCGCGCGCGCGCGCGCG 22638
Db 11027 ----- 11026
Qy 22639 CACAACCAACCCCAACACACACCTGACCTTCCCACTTCCCTTCCCAACACAC 22698
Db 11027 ----- 11026
Qy 22699 CACTTGTGCTGAATACCCAGCGCCAAACAGCGCGCGGTCAAGGCGTTCGCGCGC 22758
Db 11027 -----CCGACGCTCGGCGGCTCCCGCGCGCTCGGCGCGCGCGCGC 11068
Qy 22759 TCGGCTCCAGACACCGCGAGTTCGGGACGCGGTGAACGAGGAGACCTCCAG 22818
Db 11069 GCGCGTACGCGCGGCGCGCGGTGAGTGGGCGGAGTGCACGCGCGCGCGCGCG 11128
Qy 22819 AGCTCGCGGAAACCTCGACATGACGCTTGTCTGTGGACACGCGTGTGCGCGCATC 22878
Db 11129 GCGCGCGCGCGCGCGGTGAATGCTCCCTTCCACGCTT-----CGGCGAC 11173

Db 617 GTGAACCTCAACATCTCCGCGGAGAGCGCGTGAAGGAGGCGCTTCGGTGAATCTCC 676
Qy 12736 GCAGACGGCGGTGCAAGGGGTTTTCGGCGCGCGCGGAGCGGACCGGCTGGTCCGAGGGT 12795
Db 677 CCGAGCGCACCGCTTACACTTCGAGCGCGGGCCAAAGGATTCTGTCGGGGGAGGGC 736
Qy 12796 GTGGGAGTGTGCTGTGGTGAAGCGGCTTCGAGCGCGCGCGGAAAGGTTACCGTGTCTTG 12855
Db 737 GCGGAGTGTGCTGTACTCAAGCGGCTTCGCGGCGCTTCGCGGACCGGACCGGCTGCTCC 796
Qy 12856 GCGGTGCTCGCGGCGAGCGCGTCAACAGGAGCGGCGAGCAACGCGCTGACCGGACCC 12915
Db 797 GCGGTATCTCGCGCGAGCGCGTCAACAGGAGCGGACCGCGGCTTCAACCGTGCCC 856
Qy 12916 AACGGTGTTCACAAGTCAAGGTCAATCCGCGGAGGCTTTGGCCAAAGCAACCTCTCCCT 12975
Db 857 AGCAGGCGCGCCAGAGAGGTTGCTCGGAGGCGTACCGGAAGGCGGCTTGGACCG 916
Qy 12976 GCGGATGTGATCGGTGGAGGCGCCACGCGCAACGCGGAGCAACCTTGGGCGACCGATCGAG 13035
Db 917 TCGCGCGTCCAGTACGTGAACTCCACGCGCAACGCGGAAACCGCGTCCGCGGACCGCATCGAG 976
Qy 13036 GCTCAAGCCCTGTCGAAGCCTACGCTCAGGACGCGCGCCCAACGCGCGCCCTCTGGCTC 13095
Db 977 GCGCGCGCTCGCGCGCGCTCTCGGCTCGGCGCGCGCGCGGACGAAACCGCTGCTGCTC 1036
Qy 13096 GGAACCTCAAGTCCCAACATCGGCGCATCTCATGGCGCTGCGGGTGTGGCGGGGTCACTC 13155
Db 1037 GGCTCGGCGAAGACGATCGGCGCATCTCGAGGCGCGCGGATCTGCGGCTCATC 1096
Qy 13156 AAGATGGTATGCGCTCGGAGATGTGCTGCTCGCGGAGCGTGTGATGAGATGAGCG 13215
Db 1097 AAGACGCTCTCGCGCTCGCGCGCGGATCCCGCGAGCTCACTTCCGTACGCGC 1156
Qy 13216 TCGCGCATGTGACTGTTCCGCGGTTGCGGTGCGCTGCTGACGAGAGCGTGCCTG 13275
Db 1157 CACCGGACATCCGCTCGACACCTCTCGGCTCGAGTCCCGACGCGCTCGCGGAGTGG 1216
Qy 13276 CCGCGGGGAGGGCGGCTTACGGCGGAGGAGTGTATCATTTCCGGCTTCAGCGGACCC 13335
Db 1217 CCGACCGGAC---CGGNACTCTCGCGGCTCAGCTGTTCCGATGGCGGCGAC 1273
Qy 13336 AAGCCCAAGTATCTCGAGGAGCAACCGCGCCAAACATCCCGTTCAGACACACCGCC 13395
Db 1274 AAGCCCAAGTCTCTCAGCGAAGGCGCGCGCAG----- 1309
Qy 13396 GACGAGTCCCGGAGATCAGCGCGGAGGAGTGCCTGCTAGTGGCATGAGGCTGT 13455
Db 1310 -----GGCGGCGAGCAGCCCGGATCGATGAGGAGACCC 1345
Qy 13456 GCGCGAGTCCAGGGGTGGCGGTGGTGGTTCGGCCAAAGTCCGAGCGCGGCTTCGCG 13515
Db 1346 GTCGAGCGGGCGCATGCGCTTCGTCGTACCGCGCGGGCGGCGGCGGCTTCGCG 1405
Qy 13516 GCGAGCGGAGCGCTGCAAGCGCCACTCACGAGCAACCGCGGCTTCGACCTTCGCGGAC 13575
Db 1406 GCGAGCGCGCGCTGCAAGGCGGTGGAAGCGGACCGCGAGCTCGCGCGCGCGCA 1465
Qy 13576 GTCGGTACACCTCGCGCAACGCGCGCGGTGTTGAGACACCGCGGCAACCTTCATCGCC 13635
Db 1466 CTCGCGCGGTGCTGCTGCTACCGCGGTACCGGTCTTCAACGACGCGGTGCTGCTCGCC 1525
Qy 13636 GCGAGCGGACACCTTCGTGAAGCACTTCAGGCACTCGCGGAGCGGCGGAAACCGGACCC 13695
Db 1526 CCGGACCGCGCGCTCTTCAGCGGCTTCGGCGGCTTCGCGCGGCGGAGCGCGCGGCC 1585
Qy 13696 GCGGTATCCAGCAGACGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGGAAAGAC 13755
Db 1586 GCGGTGTTCAACGCGACCGCC-----GCCCGCGGCGGCTC 1621
Qy 13756 GATTCATGCTCGGACAGGCGACCGGACCGCGCGGATGCGCGCATGCGGCTCTACGAC 13815
Db 1622 GCGGTCTGTTGAGCGGCGAGGTTGCCAAACGTAACGCGGATGCGGATGAGGTTGTACGCC 1681

Qy 13816 ACCACCCCGTCTTCGCGCGGCACTCAACGACATCTGACCCCACTCGACCCCACTC 13875
Db 1682 GCCACCCCGCTTCGCGAGCGCTTCGAGCGCGTTCGCGCGAACTGAGACCCCTCTC 1741
Qy 13876 GACACCCCGTCTTCGCGCGGCTTCGAGCGCGCTTCGAGCGCGGAGCAACCACTC 13935
Db 1742 GACCGCGCT-----CGCGGAACTC 1762
Qy 13936 GAAGAAGCGCGCGCATGCTTCAGCAGACCGCTAGCGCGCGCGCTCTTCGCGCTC 13995
Db 1763 GTGCGCGCGGCGACACCTTCGACCGCGCTTCACACAGCGCGGCTCTTCGCGCTC 1822
Qy 13996 CAGGTGCGGCTTCACCGCGCTTCACCGAGCGCTTACCATCACCGCGGCTTACTACGCG 14055
Db 1823 GAGTTCGCGCTTCACCGCGCTCT---CGAGTCTTGGGCGTTCAGCGCGGCTCTGCTCGC 1879
Qy 14056 GAGACTCTCTCGGCGAATACCGCGCGCTTCGCGCGGCTTCGCGGCTTCGCGGCTC 14115
Db 1880 GGCCTCTCTCGCGGAGATCAGCGCGCGCTTCGCGCGGCTTCGCGGCTTCGCGGCTC 1939
Qy 14116 GCGACCGCTCATCACCGCGCGCTTCGCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14172
Db 1940 GCGCGCGCTCTCTCGCGCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 1999
Qy 14173 ATGACACCGCTTCACACCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14232
Db 2000 ATGTCGCGGCTTCGCGCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2059
Qy 14233 GACCTCGCGCTTCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14292
Db 2060 GAGCTCTCTCTCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2119
Qy 14293 ACCTGTCACACATCACCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14352
Db 2120 GCGGTCTCGAGTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2179
Qy 14353 ACCAACACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14412
Db 2180 GTCTCGAGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2239
Qy 14413 ACCAAACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14465
Db 2240 GTGGAAGCTGAGCTTCGAGGCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2299
Qy 14466 -----CCACCGGAGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14520
Db 2300 CTTGTCAAGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2359
Qy 14521 GTGGAAGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14580
Db 2360 GTACGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2419
Qy 14581 TCGGAGCGGAGCAACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14640
Db 2420 TCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2479
Qy 14641 ACCACCTTCACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14700
Db 2480 GCGAGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2539
Qy 14701 GCGAAACCGACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14760
Db 2540 CTCACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2599
Qy 14761 ACCACCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14819
Db 2600 GCGAGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2659
Qy 14820 ACCACCGGAGCGGAGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 14879
Db 2660 GCGCGCGCGGAGCGGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTC 2719

Db 17452 CCCGTCGGGGGTGCGGGCGATGGATCCCGAGCGTGTGTGGCGGTGATGGCTGATGCGG 17511
 Qy 24245 GGGCGAGCGCGCGAATTCACAACTGATCATCGCGACATCGACTGGAGCGCTTCGTCC 24304
 Db 17512 TGGGTGCTGTGAGCGGTTCGTGCG---GTGCTGATGTGGACTGGGAACGTTTCGTCA 17568
 Qy 24305 CCGCTTACCGCTGAGAGCCACAGCCCGCTCATCTGAGGACATTCGGAGGTTTCGGCAAG 24364
 Db 17569 CCGGTTTCGCTTCTGCGCGTCCCGTCCGTGTGATCATGATGACCTCCCGAGGTGCGTGTG 17628
 Qy 24365 CGCTCAGGAGCTGGAAGACGCTGCTGCGACGCAAGACGACCAAGCTCAGCCGATG 24424
 Db 17629 TTGTGAGGCCAGGTTCAGGGCCGCGGCGAGGGTTGGCTTGGTGTGGAGGAGT 17688
 Qy 24425 CGAGCTCTCTCCGTGAGCGAATGGCCCGACTGACGCTCTCAAGCAGAACCAAGTGTCTG 24484
 Db 17689 CGTCGGGTGGTTGAGCGGTTCGTGGGGTTGTCTGTGTGCGCAGGAGGAGTTGG 17748
 Qy 24485 TCGGCTGATTCGGACAGGCAATCTGCACCGTTCTCGGCTTCGTAAATCCGGAAGGCATCG 24544
 Db 17749 TGGAGTTGGTCCGTGCTAGGCTGCCGTTCCTCTCGGGCATGGTTCCGCGCAGGACGTC 17808
 Qy 24545 AGGACCAACAGGCTTCGGGACCTCGGCTTCGACTGCTGAGCTGGCTCAGTTACGCA 24604
 Db 17809 CGGCTGAGCGGGCGTTCAAGGAGTTGGTTTGAATCCCTCACTGTGTGAGCTACGCA 17868
 Qy 24605 AGAACTCGCAAGGAAACCGGACTGCACTCCCCCGCTCCCTGGCTTCGACTATCCCA 24664
 Db 17869 ACGGCTGGCCGCGGCACCGGATCGGCTGCGGCCACCATGGCATTCGATCATCCCA 17928
 Qy 24665 CCCCAGGAAATGTGTGCCCATCTCGGCAACAACACTCGT----- 24704
 Db 17929 CCGCCACCGCATCGCACGCTTCCTGCAATCCGAATCGTGGGAAGTACGACCGCTGA 17988
 Qy 24705 ----- 24704
 Db 17989 CGCTCATGCGGTGCGGCATCGACAGTTGGAGACCGTCTGGCTCTGCTGGAATCGGACG 18048
 Qy 24705 ----- 24704
 Db 18049 AGAAGCTCGCTCGGAATCACGAAGGATTGAACATTCTTCTGCCCGCTTCGGAAGCG 18108
 Qy 24705 ----- 24704
 Db 18109 GAGGCAATTCGAGAGCAGGAGAGCAGGACAGACAGCGGAAACATCAGGATGTCGAGG 18168
 Qy 24705 ----- 24704
 Db 18169 ACGCCACCATCGATGAGCTATTTCGAGGTGCTCGAACGAACCTCGGCAATTCCTGAAAC 18228
 Qy 24705 -----CGACCTAGACGACG 24718
 Db 18229 CTGTCCGACTGTACCGGACCTTGACCGGAGACGCTGTGAGAACGACGAAAGATCG 18288
 Qy 24719 AAGAGGACCGGACATGTGGAATGTCTCCGCAAGTGGCCCATCGGCGTACC-----G 24772
 Db 18289 TCGAGTATCTCAAGCGCGGACCGTGGACCTCGGCAAGGCCCGGCACCGCATCTGGGAC 18348
 Qy 24773 TGAGAGCAAAACGATCGGCATCATCTGATGATGATGCTTCCCGCGGCGGTACGTT 24832
 Db 18349 TGAGAGCAGGCCATCGCGATCACGTGATGGCTTGCCACTTCCCGGCGGATCGAGA 18408
 Qy 24833 CTGCGACGACATGTGGGAATGCTCGCTTCGGTAAAGACGCTATCGGCGCTTCCCGA 24892
 Db 18409 GTCCGAGCAGCTGTGGGAATCTCTTCCGCGGAGCGAGGTGCTTCCGAGTTTCCCG 18468
 Qy 24893 CCGACCGGCTGGGACCTGGACACGCTCTTACGACCCCGACCCCGACCAACCCCGGACCT 24952
 Db 18469 ACGACCGGCTGGGACCTGGACGAGATCTACCATCTTACCCGCGGAACACAGTGGGACGA 18528
 Qy 24953 GCTACACCGGAACGGCGATTCCTTACGGCGAGGCGCACTTCGAGCGCGGAATTCG 25012
 Db 18529 GCTACGTCGCTCACGGCGGTTTCTTGGATCATGCGACGAGGATTCGACACGAGCTTCTTCG 18588

Qy 25013 GCATCAGCCCCCGGAGGCCCTCGCAATGGAACCCCGAGCAACGACTCTCTCTCGAAACCG 25072
 Db 18589 GTATCTCGCGGTGAGGCGTTGGCGATGACCCCGCAGCAGCGGTTGCTGCTGAGACGT 18648
 Qy 25073 CTTGGGAAACCATCGAACAACGCGCGGATCAACCCCGACACCTTCACGGGACCCCGACCG 25132
 Db 18649 CTTGGCAGCTTTTCGAGCGCGGAGGTGATCCCCATACGTTGAAAGGAAGCGGACCG 18708
 Qy 25133 GAGTCTTCCCGGAAATCAACGCTCAGACACGCGCGCATATCCGCCAAAGCCGATG 25192
 Db 18709 GAGTATTCGTCGG-----CGCCGCACACATGGTTATCGGACAGGGTGAACCTCCGC 18762
 Qy 25193 TGGAGACCATCGAGGCTACGCGCTGACCGCGAGTTTCGGGAAGTGTGGGCTCCCGCGCGG 25252
 Db 18763 CGCGGAGCGCGAGGGCTACCTGCTCAGAGGAAACGCTCGGCGGTGTCTCGGGGTA 18822
 Qy 25253 TGGCTTACAGCTTCGGGCTCGAAGGCCCGCGGTTGCTGTGATACGCGGTGTTCTGCT 25312
 Db 18823 TTTCTTACACTTCGCGCTTGGGGCTTGGGCTGACGCTGACGACGCGGTGCTCGTCT 18882
 Qy 25313 CGTTGGTGGCTTGCATTTGGGCGCGCAGGCGTTGGTCCGGTGTGATGTTGATGCGCG 25372
 Db 18883 CGCTGTGGCTGCTGACCTGGCGGTGACGCGCTGCGCGTGTGGCGAGTGTCTGCTGGCGG 18942
 Qy 25373 TTGCGGGGTGTGACGCTGATGTCTCTCCGGTACGTTTGTGGAGTTCTCACGTCAGC 25432
 Db 18943 TCGTGGTGTGTGGCGCTCATGTGCGACCCGAAAGTCTCTGCTGAGTTTACGCGGCGC 19002
 Qy 25433 GGGGTCTGGCCCGGACCGGCGGTGCAAGGCTATTTGGCGCTGTGTGACGTTACCGGT 25492
 Db 19003 CGGACTGGCCAGGACCGCGGTCCAAGGCTTTTGGCGCTCAGCGATGTTTGGGT 19062
 Qy 25493 GGGCGAGGTTGGGATGCTGCTGGTGGAGCGGCTCTCCGACCGCGCTCGCAACGCTC 25552
 Db 19063 TCGCGAGGAGTTTCTGCTGCTTCTGCTGGAGCGGTGTGCGATGCGCGCGGTGGGTG 19122
 Qy 25553 ACCGTGCTCTGCGCGTGGTGGCGAGTGGCGTCAACAGGACGCTGCGAGCAACGCTC 25612
 Db 19123 ATCGGCTGTGGCGGTGGTGGCGGAGTGGCGTCAATCAGATGTTGCTGCTGCTG 19182
 Qy 25613 TGACCGCGCCAAACCGGCGCTCCAGCAGCGTGTATCTGTCAGGCTTGGGCAATGCGG 25672
 Db 19183 TGGCGCGCGCAATGCTCGCTCGCAGCAGCGGTGATTCGTGCGGCTTGGCTGACGCGG 19242
 Qy 25673 GACTGACCCCGCGCTGTCGACGAGTGGAGGCGCAACGCGACCGGACCTCTCTGGGG 25732
 Db 19243 GTCTGGCTCTCCCATGTGGATGTGGTGGAGCGCATGGTACGGGAGCGCGTTGGGTG 19302
 Qy 25733 ACCCGATCGAGCCCGCAGGCACTCTTGGCGCGCTACGGAACAACACCGCCCGCCACCGCC 25792
 Db 19303 ATCCGATCGAGGCTCAGCGGTGCTTGGCGAGTATGGGCGGGGCGTACAGTGGGCGTC 19362
 Qy 25793 CTTTGGCTGGGATCCCTCAAAATCGGCAACGCAAGCGCGCGCGCGGCGTGG 25852
 Db 19363 CGGTGTGGCTGGGTCGTGAAATCGAATCGATCGATCAGCAGCGCGCGCGGCTGCG 19422
 Qy 25853 GCGGAGTATCAGATGTTGATGGCCCTTGGCAACGCGGCTGCTGCAAGACCTCTCCAG 25912
 Db 19423 CTGCTGTATGAAGATGTTGCTGGCTCTGGAGCGGGGTGTGGTCCGGAAGAGCTTGCACG 19482
 Qy 25913 TGGAGAGCCCAACCCCGCAGGTCGACTGTGTCACAGGCGAGTACAACTCTCTGACACAAC 25972
 Db 19483 TGGATGAGCGCTCTCCGATGTGGACTGTGATCGACCGGTGCGGTGAGTTGCTGATGAAG 19542
 Qy 25973 CGGTGCTTGGCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 26032
 Db 19543 AGCGCGCGTGGAGCGCGA---GGCTGAGCGCTTCTGCTGGGCGAGGCAATTTCCGCGCTTCG 19599
 Qy 26033 GCGTCAGCGGCGCAACCGCCCATCATCTCTGGAAGAGACCCCACTCCCGCAGCAGCG 26092
 Db 19600 GTGTCGTGGCAAGAAATCGCATGTGATGCTGTGGAGGAGGCACTCTCGGAACCGGACCGG 19659

QY	22040	TATGCAAAACCATGCCCCCGGACCATGACCAACCTCCACACCAACCCACACCAATCA	22099	Db	16402	ACCGCC-----GGGCTTATGCGGAGGCTGTGCGGGCGGTGTTGAGCGGG	16448
Db	15352	TGCCCCGATCGCTGTGTGGGGCGGATGCTCTCGGTCTTTTTCAGCTGAGGTGTC	15411	QY	23165	CGGACCATCATCCGGCTGTCTCTCCCTCTGCTCGCCCTCGAGAAACACCCACCCCAACC	23224
QY	22100	CCACACCATCTACCGCCGCAAGAAACGACTCGCCATCGCGGCATCAACACCCCACT	22159	Db	16449	GTCTGTGTCTTCTTGTCTGAGATGATCGGGGCACTCGGAGCATTTCTGTTTCCCGGC	16508
Db	15412	GCACCATGTCTGACACACTACGGCGGAGGGTTTTCGTCTCGCGGCTCAATGCGCCGCTCT	15471	QY	23225	ACCCACACACACCAACCGGCACCCCTCTCAACCTCACCTCCCTCCCAACCCACACCAAA	23284
QY	22160	CCCTGTCTATAGCGGCACCCCCACACCGTCCAAACACATCACACCTCTGCCAACAC	22219	Db	16509	GGTCTTTCGGCGTCTGCTGTTTGGCGCAGGGT---TGGTTGATCTTGGCCGGTTGGTG	16566
Db	15472	CGACCTGTGTCTCGGTGAGCCCGAGCCCTGATGATGTTGTTGGCCGGTTGTAGCGGG	15531	QY	23285	CCACACCAACACCCCTCTGTTAGCCACACCAACCCACCAACCCACCAACG	23344
QY	22220	AAGGCATCAAAACCAAAACCTCTCCCAACCAACCGCTTCCACTCCCCCAACCAACC	22279	Db	16567	AGGGGCGCGGTGTGCTGTGTTGACGCGGGGTGCGTGTGCTGCTCTCGATGCGG	16626
Db	15532	AGGGTCTCGGGCTCGTGTGTTCCCGTGTGACTATGCTCCCACTCCGCGCAGATGAGC	15591	QY	23345	ACCCCTCACACACCCCAACCCCAAGCCAAACCTGGGGACTCGCCGCAACCACTCTCTG	23404
QY	22280	CATCTCTAACCACTCTACACGACACACCAACCTCTACCTACGACCAACCCACACC	22339	Db	16627	GTGTGTGATGATCGGTGACGGCGAGGTGTGGGGTTTGGGGTGTCTTGGGTCTGG	16686
Db	15592	AGTTACCGATGATGTTGCTGGAGGCGCTGGCGGACGTCACTCGCAGGACTCCAGTGTTC	15651	QY	23405	AAACCCCAACCAACCCCGCGGAATCATCGACCTCCCAACCAACCCCAACCAACC	23464
QY	22340	CCCTCATACCGCAACACCCCAACCGACCAACTCTCAACCCCAAC-----TACT	22390	Db	16687	AGCATCCCGAGTTGTGGGTGGGTGTGACCTGCGGTGGGGTGTGATGAGGAGTGT	16746
Db	15652	CGTTTTTCTGACGGTGAACCGGGAATGCGTGTGACGACACCGCTCTGGATGGGGTACT	15711	QY	23465	TCCACCACTTACCAACCCCTTCAACCAACCCCAACCAACCAACCTCTGCGCATCGCA	23524
QY	22391	GGACCAACCAAGCCCGCAACACCTGCTGACTAGCCACCAACCAACCCCTTCCACCAAC	22450	Db	16747	GCGCGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTG	16806
Db	15712	GGTTACAGAACTGCGGGAGACGTTCCGGTTCAGGAAGCGTTGAAGGCTTGTGGCTC	15771	QY	23525	CCACCGGCACCAACCCCGCGCTCAACCCCAACCAACCTTACCCCAACCAACCAAC	23584
QY	22451	ACGGGTACCACTATCTGAACTCGGACCGGACCAACCAACCTTACCAACCTTACCCACC	22510	Db	16807	GTTCGGGTGTGTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTG	16860
Db	15772	AGGGGATGGCGGCTTCTGCTGAGTGAGCGCGCACCTGTCTCTGCTCCGGGATCACAG	15831	QY	23585	CCACCCCAACCCCAACCAACCTTCAACCCCGGGAACCGCGCTCTGCGCACCC	23644
QY	22511	ACAACCTCCCAACACCCCAACCAACCTTCAACCTTCAACCTTCAACCTTCAACCTT	22564	Db	16861	GTGGTTGGCGCGGTGGAGCGGTGTGTTGTGATCGGGTGTGTTGGTGTGTTGGTGTG	16920
Db	15832	AAACACTCGACACTTGGAGCGCGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	15891	QY	23645	ACCTCACCAACCACTTCAACCAACCAACCAACCAACCAACCTTCTCTTCAACCGCA	23704
QY	22565	ACCCCAACCAACCTTCTTCAACCACTTCCCAACCAACCAACCAACCTTCCCAACCA	22624	Db	16921	ATACCGCCCGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGG	16980
Db	15892	GCGGCTGTGATCGTTTCTTCACTGCTTCCGAGGCTTCCGAGGCTTCCGAGGCTTCCG	15951	QY	23705	CGGCGCCCAACCAACCCCAACCAACCTTCAACCAACCAACCTTCAACCAACCAACCA	23764
QY	22625	ACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCTTCCCAACCTAC	22684	Db	16981	GTGGCAGTGGCGCTGCTGGGTGTGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGG	17040
Db	15952	ACTGGACCAATGCTTCTGAGGGTGGAGCGCGCG---GCTTGTGCTGACCTTCCCACTATG	16008	QY	23765	ACCTCACCACTTCAACCAACCAACCAACCAACCAACCTTCAACCAACCAACCTTCA	23824
QY	22685	CTTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCTTCCCAACCT	22744	Db	17041	GGGTGTGTTGGCGCGCTGTGATGTTGGGTGTGTTGGGTGTGTTGGGTGTGTTGG	17095
Db	16009	CCTTCCAGCAACGCTTACTGCT---GCAGAGAGCGCTGCAAGCGGTGCTGA-16063	16063	QY	23825	CCATCCCAACCAACCAACCTTCAACCAACCAACCAACCAACCTTCAACCTTCTCTG	23884
QY	22745	GGGTTCTCGCGCTCGGCTCCAGACACCGCGAGTCTGGAGTTCTGGGACGCGGTGAACG	22804	Db	17096	-GTCGATCTGGGTGAGCGGTGACGCGGTGTTCCATCGGCTGTTCTCTCAGTCA	17154
Db	16064	-----TGAGCGGTGGATGCGGAGTTCTGTTCTGTTGTTGGAAC	16101	QY	23885	CCCCCGTTCGGAACCGATGCCGAATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	23944
QY	22805	AGGAAGCTCCAGAGCTCGCGAAACCTCGACATCGACGCTCTGCTTGGACACGG	22864	Db	17155	CGCCTTGGCGGAGATCTCTGTCTAGGAGGCGGTGATGTTGATGGCGGCAAGGTGGCG	17214
Db	16102	GCGGCGATGCAAGCGGTGTCGATTTGCTGAGCAGCGCGCGGCTTTCACACCG	16161	QY	23945	GGCGCGCATCTGTCATGAGTTGCTGTCGACCATGAACGCTTGAACACTTCTTCTCT	24004
QY	22865	TGTCGCGCAGCTCTCGCTGCGCAGCGCCACCAACCAACCAACCGCGCATCAACCT	22924	Db	17215	GTGCGGTGAATCTGGGTGAGTTGGTGGATCCCTGTG---GTCGAGGCGTTTGTGTTGT	17271
Db	16162	TGTTGCGGCTTGTGTTGTTGCGGGGGGTGCGGTGAGCATCGACGGCTTCAAGACT	16221	QY	24005	TCTGTCTGGGCGCGCGCTTGGGCGAGCGGAATCTAGTGGCATCTCTGCGGCGCAACG	24064
QY	22925	GGACCTACAGGAACCTTGAACCCCTCAACCTTCCCAACCAACCAACCAACCAACCA	22984	Db	17272	TCTCTCTCAATGCGCGGTGTGGGCGAGTGGGGGCGAGCGGTGTATGCGGCGGCAATG	17331
Db	16222	GGGTTTACCGGGTGGAGTGGAGCCCTTCCCGGCGCGCTTGTATGAGTGTCTGCTGGTGTG	16281	QY	24065	CATACCTGAGCGCTTCCGAGCATCTGTCAGACATGGAATTTCCCGGGCATCGATCG	24124
QY	22985	CTGGCTCATCGCCATCCCGAAACCCAGACCAACCAACCAACCAACCAACCTTCTCA	23044	Db	17332	CGTTTCTGATGCGTTGGCGGTGCGTGTGCGGGGTGTTGGTCTGCGCGCCAGTGTGG	17391
Db	16282	GCTGTGTGTGTTGTTGTTGCGCGGGCTTGGCGGATGATGTTGTTGTTGCGGGTGTG	16341	QY	24125	CTGGGCGCTTGGGCGGAGGATGTGCGCGCGGTGATGCGCGCTCATGTTTACCTCG	24184
QY	23045	CCAACCTTCAACCAACCGCATCAACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCT	23104	Db	17392	CGTGGGAGTGTGGGCTGGTGGGGATGGCGTCTGGTGGGTGGTGGGCGGGAGTGT	17451
Db	16342	CTGCGGTCAAGCGCGGGGTGGGAGGTGAGTGTGTTGAGTCTGATCCGACCCCTCTG	16401	QY	24185	AAAAGCGGCGCATCTGCGGATGGAGCCACGATGCGCGCTCGCGGATTTCCATCTGTGCG	24244
QY	23105	ACCCCAACCACTTCAACCAACCTTCCACCAACCAACCAACCAACCAACCAACCAACCA	23164				

Db 13222 TCAGCTCGCTGACCTCGGTGCGAATCGCGCAACCGGCTGAACACCGCCACCGGCTCAGAC 13281
QY 19883 TGCCGACGACTCTCGCTTCGTATTACCCAGCCACCAAGCTCGCCCAATTCTCTCT 19942
Db 13282 TGCCCGTACGCGCGCTTCGTACTAGCGAGGCCCGGGCGCTGGCGGCCACTCTCGCT 13341
QY 19943 CCGAGATCGCGGAGTTTCAGCCCGCAGCAACTCAACTCCGCTTCGCGACCCCGGCGAGAGC 20002
Db 13342 CCAGGCTGATCGACGACGATGGTGACACAGTGCCTTCCCGCGCTGGAGAACGACGCGA 13401
QY 20003 TCGATGAGCGGATCGGCATCGTTGGCATGGCCCTGTCTCCCGCGGAGTGAACCTCGG 20062
Db 13402 TCAGAGCGGATCGCGGATCGTGGAAATGSCATGCGCGCTTCGCGGAGGATCGCTTCCC 13461
QY 20063 CGGACGACTTCGCGGATCTGATCTCTCGGAGCAGGACGCGATCGCGGATTCGCCACCG 20122
Db 13462 CGGAGGATCTGTGGGAGCTGCTCACCCTGGTGAGGACGTTGTCTCGGACTCCGCCAGA 13521
QY 20123 ACCCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCCCGGCACTGCT 20182
Db 13522 ACCCGGCTGGGACTTGGGCGCTGTAGATCCCGATCGGACCGGCGGTACGTCAT 13581
QY 20183 ACACCGGACGCGGATTCCTCTACGACGCGGCACTTCGACGCGGAAATCTTCGGCA 20242
Db 13582 ACATGCTGAGGCTGCTTCTCTGACGAGCGGGGAGTTGACGCGGCGCTTCTTCGGTA 13641
QY 20243 TCAGCCCGCGAAGCCCTCGCCATGAGCCCGCCAGCAACGACTCTCTCGAAACCGCT 20302
Db 13642 TCTCGCGGTGAGGCTTGGCGATGAGACCCGCGAGCGGTTGCTGTGGAGACGTCCT 13701
QY 20303 GGGAAACCATCGAAACGCGCGGATCAACCCCAACCCCTTCAACGCGACCCCGGAG 20362
Db 13702 GGGAGGCCCTCGAAACGCGCGGATCACTCTTCAAGCTGGCGGCGAGTCCGACCGGTG 13761
QY 20363 TCTTACCGGCAACGAGGACGAGGACCGCGGCAACATCGTCAAGCCCGCGAGCGGTA 20422
Db 13762 TGTTCCTCGGCAATCGAAACGAGGACTACGCGCGCGGCGAGCTGCGCTTCGAGC 13821
QY 20423 CGGAGGATTCGTCTACCGCGGAGCAGCAGCATCGCTCCGCGCAATCTCTACA 20482
Db 13822 TGGAGGCTACCTGTCTACCGGCTCATCTCAGCGTCTGTTGGGCGGTGTGCTTACA 13881
QY 20483 TCTCGGTTGGAAGGCGCTCGGTCACCTCGACACAGCTGTCTCTCTCGCTCGTCG 20542
Db 13882 CGTTCGCTTCTGAGGCGCTCGGTCACGCTGATACGCGCTGCTCGTCTGTTGGTG 13941
QY 20543 CCTGCACTCGCTCGGCTCGCTCAGTTCGCTGAGTTCGATGACGATGCGCTTGGCGGCG 20602
Db 13942 CGTTGATCTGGCGGTGCAAGGCTTGGCGGCGGTGAGTGTCTGCTTGGCTTGGTGGGTG 14001
QY 20603 GGGCACGCTCATGACACCCCGATCACCTTACCGAAATTCGCGCGCCACGCGGACTCG 20662
Db 14002 GGGTACGCTGATGCTCGCGGTGACGTTGACGAGCTTCTAGTCGGGACGCGGTTTGT 14061
QY 20663 CCCCGACGCGGCTTTCGAGGCGTCTCGCGCGCGCTGACGCTACCGCTCGGCTGAGG 20722
Db 14062 CGGTGATCGGCGGTTCGAGGCGTTCGCGCTTCGCGGATGCTTGTGCTGCGGAG 14121
QY 20723 GTGTGGGATGCTGTGTGTGGAGCGGCTTCGACGCGCGCGGCAACGCTACCGGTCTCC 20782
Db 14122 GTGTGGGTGTGTGTGTGGAGCGGTTGTGCGATCGCGGCGGTTGGGTTCATCGGGTGT 14181
QY 20783 TGGCGGTGTGTGTGTGAGTGGGTCAACGAGGCGGTTCGAGCAACGCTCTGACCGCGC 20842
Db 14182 TGGCGGTGTGTGTGTGGGAGTGGGTCAATCAGATGCTGCTCAATGCTTGGCGGCGC 14241
QY 20843 CCAAGCGGCTTCCAGGACGCGTCAATCCCGAGGCGCTTCGCGCAACGCGGACTTGACCC 20902
Db 14242 CGAATGCTCGTCCGACGAGCGGTGATCCGTGCGGCGTTCGCTGACGCGGCTTGGCTC 14301
QY 20903 CGCGGACGCTGATGCGGTGGAGGCGGACCGGACCGGCAACGCTTGGCGGACCGGATCG 20962
Db 14302 CTGCGGATGTGATGTGTGGAGGCGCATGGCACGGGGACGCGGTTGGGTGATCCGATCG 14361

QY 20963 AGGCCCAGGCCATCTCGGACCTACGACAGGACCGTCCCGCAACGCGCCGTTGTGGC 21022
Db 14362 AGGCTCAGGCGTTGCTGGCGACGATATGGCAGGGTCGTACCACTGGGCGTCCGGTGTGC 14421
QY 21023 TGGGCTCGTCAAGTCCAAAGTCGACACACACAGCCCGCGGCGGCTGGCGGAGTGA 21082
Db 14422 TGGGCTCGTCAAGTCCAAAGTCGACACACACAGCCCGCGGCGGCTGGCTGTGTGA 14481
QY 21083 TCAAGATGTGATGCGCTCCCGACCGGACACTCCACCGACTCTCCACGCGGATGAGC 21142
Db 14482 TGAAGATGTGCTGGCGTTGGGTTCGGGCTGGTCCGAAGACGTTGCTATGTGATGAGC 14541
QY 21143 CGTCCCGCATCTGCACTGCTCCCGGCTGGGTGAGTGTGAGTGTGAGGAGCGTCCCT 21202
Db 14542 CGTCCCGCATCTGCACTGCTCCCGGCTGGGTGAGTGTGAGTGTGAGGAGCGTCCCT 14601
QY 21203 GCGCGGCGGAGGCGGCGCGCGGCGGAGTGTGATCATATTCGCGGCTCAGCGGCA 21262
Db 14602 GGTCCGCGG---TGGCGGCGGTGCGCGGCGTGGTGTGTGCTGCTTCGGGATCAGCGGCA 14658
QY 21263 CCAACGCGGACGTCATCTCGGAAGACACCGCGGACGAGCTTCGCGGCGGACCGG 21322
Db 14659 CGAATGCGCATGTGATGCTGGAGGAGGCGCTTCGCGA----- 14695
QY 21323 CCGACGAGATCCGCTAGTGGCGGAGGCTGCTCCCGGAGTCTCTGGGCTGTGGCGCT 21382
Db 14696 -----GCTTCGTTGGTGGAGGAGGCTCGGCTCCGTTGTGGTGTGGTGCCT 14742
QY 21383 GGTGCTGTGCGGCAAGTTCGAGCGGCGCTCGCGCGGCGGAGGCTTCGCTGCTGAGCGCC 21442
Db 14743 GGTGCTGTGCGGCGGAGTTCGCGGCGGCTTCGCTGCGACAGCGGCACTGCTGGCTGCGC 14802
QY 21443 ACCTCACGACCAACCGCGGCTCGACCTCGCGAGTCCGATACACCTCGCGGCGGCG 21502
Db 14803 AGCTGTC---GAGCACGCGGTGCGGCTGCTGATGCTGGGCTGGTCTGGTGGCGCACGA 14859
QY 21503 GCGCGCTGTTCGACCAACCGCGGCACTTCATCGCGCGGCGGAGCGGACGCTTCCTCGAAG 21562
Db 14860 GGTGCTGTTCGAGCAACCGGCGGCTAATGCTCGGCACTGATCTTGATTCATGGCGGGT 14919
QY 21563 CACTCCAGGCACTTCGCGGAGCGGCAACCGCGCGGCTCATCACAGAGCGGCGGCG 21622
Db 14920 CGTTCGCGGCTTCGCTCGGCTGCTGTCGCGGCGGCTGCTGCGGCTGGGCTCGCG 14979
QY 21623 GCGGACCGGAGCGGCGGAGCGGAGGAGGACCGGCAACGCTGCTCCGAGCAGGCGCA 21682
Db 14980 CTGAGGCGCTGCTGTGCTG---TTGCTCTTCTTCCTGCTGAGGGT 15021
QY 21683 CCCAACGCGCGGCTTCGCGGCTTCACGACACCGCGGCTTCGCGCGGCGGCGAC 21742
Db 15022 CGGAGTGGGTGGGATGGCGGCTGGTTCGATGCTGCTGCTGCTGCTGCGGAGCGG 15081
QY 21743 TCAACGACATTCGACCCACCTTCGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 21802
Db 15082 TGGCGGAGTGTGCGCGGCTGCTGAGCGGCTTACCGGCTGGCTGCTGAGGCTGTT-- 15139
QY 21803 CCCAAAACGACAAACGACGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21862
Db 15140 -----GCGTGGTGGTGAAGCTGCTTCTTGGGCGGGTGTGATG 15174
QY 21863 AGCGCGGCGGCGGCTTCGCGCTTCGAGGCTCGCGCTCCACCGGCTCTTCACCGGCT 21922
Db 15175 TGTGCGAGCGGCGGTTGTGGGCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 15231
QY 21923 ACCACATCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21982
Db 15232 AGGCTGTGGAGCTGCTGCGGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15291
QY 21983 TCGCGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 22039
Db 15292 TGGCTGGGCGGTTGAGTCTGCGCGGATGCTGCGCGGCTGCTGCTGCTGCTGCTGCTG 15351

QY 17702 CCACCACCAACCAACCTTCCACCAACACGGGGTCAACACCTACATCGAACTCGGACCG 17761
DB 11263 AAGAAGCCGTGGAAGGGCTCGTGGCTCAGGGATGGCGGTGCTGTCGAGTGCAGCCCG 11322
QY 17762 ACACACCTCAGCACCTTACACCAACCAACCTTCCCAACACCCCAACCAACCTTCA 17821
DB 11323 ACCCGCTCTCGTCCCGGGCATCAGAAACATCGACACCTTCGACGCGCGCTGTG 11382
QY 17822 CCTCACCACACCCG-----ACCACACCCCAACACCTTCCCTCACCACCTCGCCA 17875
DB 11383 CACTGTGCTGTGCGGGGTGAGAGGGGCTGTGATCGTTCTTCACTCCTCTCGGG 11442
QY 17876 AAACACCAACCACTTGGCAACCCCACTTACACCAACCAACCAACCAACCAACCC 17935
DB 11443 AAGCTTGTTCAGGCGCTCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11502
QY 17936 ACACCACTTCACTTCCCACTTACCTTCCCAACACCAACCAACCAACCAACCAAC 17995
DB 11503 GCACCG---TCGACCTTGGCCACCTTACCTTCCCAACCAACCAACCAACCAAC 11559
QY 17996 CACAGCGGGTCCGGATCGGTTCGGGTTCCGGTTCCGGGGCGGCGAGGACTGCGGG 18055
DB 11560 AGCGGGG----- 11567
QY 18056 GACGCGCAGAGGTGAGTGGCGGTCTTGGGACGCGGTGGCGCGCCAGGACCTGGAAC 18115
DB 11568 ----- 11567
QY 18116 TCGGACCACTGCGCGGTGCGCGCTTCCCGCGCGCTGACACGAGTGGTGGCGGACT 18175
DB 11568 ----- 11567
QY 18176 CCGCTTGGCACCGCACCAACGACGACGACGACGACGACGACGACGACGACGACG 18235
DB 11568 ----- 11610
QY 18236 CTTGGAACCTTCACTTCCCAACCAACCAACCAACCAACCAACCAACCAACCAAC 18295
DB 11611 AGTGGCGCGCTCAGCAGACGACCTTCCGACGCGCTGCTGGCGGTGGCGA 11670
QY 18296 TCCCGGAACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 18355
DB 11671 TCCCGGACGCTTGGCGCGGTGACTCATGTTGGTGGCGGCACTCGACGACGAG 11730
QY 18356 AGCGCATACCCCTTCCCTTCACTTCAACCAACCAACCAACCAACCAACCAACCT 18415
DB 11731 GCGCGCGCTGCGGTGCGGTGCTGATGAGCGGACGACGACGACGACGACGACG 11790
QY 18416 ACCACCGGACCAACGACCAACCAACCAACCAACCAACCAACCAACCAACCTTCC 18475
DB 11791 TCGAGGCTTTCGGGACGCGCTGGGCGAGACG-----ACCTCGCGGTGCTCTCC 11844
QY 18476 TCTCGGCTTGAAGAAACACCCACCGGACGACGACGACGACGACGACGACGACGAC 18535
DB 11845 TTTTGGCGCTTGAAGCGCGGACGACGACGACGACGACGACGACGACGACGACGAC 11904
QY 18536 TCAACCTTCACTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCTTCT 18595
DB 11905 CTTTGGCTTGGGCTTGGTGGGCTGATGGCGACGCGCGGCGGAGGTGCGGCTGAT 11964
QY 18596 CACCAACCAACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 18655
DB 11965 CGACCGGCGCGGTGAGCGGCACTGCGCGGTGATCAGCGGACGACGACGACGACG 12024
QY 18656 AATCTTGGGACTTGGCGGACGACGACGACGACGACGACGACGACGACGACGACG 18715
DB 12025 TGCTTGGGACTTGGGAGCGGTGCTGCGCTGGAACAGCGGACGATATGGGCGGTG 12084
QY 18716 TCGACCTTCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCTTCA 18775
DB 12085 TCGACCTTCCCGGAACCTTGGACGAGACGCGGGCGACGCGGTGCTGATGTGTTGG 12144
QY 18776 AACCCCAACCAACCAACCAACCTCGCATCCGACCAACCAACCAACCAACCAAC 18835

DB 12145 GCCTGGCGGACGAGGACCACTTGGCGGTACGGGCTCTCTCGCTCTCGCGGACGCTCG 12204
QY 18836 CCCCACCAACCTTCAACCCCAACCAAC-----CACCCACCCCAACCCCAACG 18886
DB 12205 TTTGATGCCCGGTACCGTATGTGAGGACAGGGCGGGGCGGAGTGTGTCGCCACG 12264
QY 18887 GAAACACCTTCACTCAGCGGCGGAAACCGGCGCTTCCGACCAACCTTCAACCAACCTCA 18946
DB 12265 GACGGTCTGTGTGACCGGAGGACCGGGGCGCTGGGCGGACGTCGCGCTGGCTGG 12324
QY 18947 CCACCAACCAACCAACCAACCTTCTCTTCAACGAGCGAAGCGGCGGCGGCGGCGG 19006
DB 12325 CCGCAAGG---CGCGGAGCACTGTGTAATCATCAGCGCTCGGAGGACGACGAGCG 12381
QY 19007 ACGCACAACCTTCAACCAACCAACCTTCAACCAACCAACCAACCTTCAACCACT 19066
DB 12382 GGGCGCTGCTTGGGACAGCTTCAAGGATGAGGTGTGCGGTGACCTTGGCGGCT 12441
QY 19067 GCGACCAACCAACCAACCAACCTTCAACCAACCTTCAACCAACCAACCTTCAACCA 19126
DB 12442 GCGATGACGAGGACCGGACGACCTTGGAGACGCTCTCTGACTGCTGCGACGATCG 12501
QY 19127 CC---CTCACACCGTCTATTCACACCGCAGGATCTCTGAGACGCGCACTTCAAC 19183
DB 12502 CCGAGCTGACGCGCTCATCCACGCGCGGTGCTCTGAGACGCTCTCTGAGACGATG 12561
QY 19184 TCACCCCAACCACTTCAACCAACCTTCCGCGCAAGCCACACGCGCCACCTCTCC 19243
DB 12562 TCACACCGGACGATGAAACCAACCGCTTGGAGGAAAGTACGCGCACCTTCAAC 12621
QY 19244 ACCAACTCACCAACCAACCAACCTTCAACGCTTCTCTCTCTCTCTCTCTCTCT 19303
DB 12622 AGAACTGACCGGAGCTTCAACCTTCTCGGCTTCTGTAATCTCTCTCTCTCTCT 12681
QY 19304 CTTTGGCGGACCGCGGCAACCACTTACGCGGAGCAAGGCTTCTCTGAGCGCTCTG 19363
DB 12682 CCGTGGAAATCCCGGCGGAGCAACTACGCGCGGAGAACTCTGTTTGGAGCGCT 12741
QY 19364 CCCACACGCGCACACCAACCACTTCCCGCCACCAACGCTTCTGCGGACCTTCTG 19423
DB 12742 CGGAATGCGGAGGCTCAGGGCTCTGCGGACCTTCCATCGCTTGGGACGCTG 12801
QY 19424 GAAACGAGCTGCGGACTGCTCAA-----GTGAGCAACATCTCCGCGCGCGG 19477
DB 12802 GCGGACCGGATGACATGAGGCTGCTGGGCGAAGCTTCCAGCGGACGCTGTAC 12861
QY 19478 TCGCATCGCGCGCGGCTGAGTGGCGTCAAGCTGTTGACGCGGCGATTCGAGCGGCG 19537
DB 12862 TCGCATGGAACCGCGCGGCGCTGCTGGCTCGACCAACGCTTGGCGAGCGA 12921
QY 19538 CGAGTCTCTCTGTCGCGGATTCGACTGGAAGAAATGGGACCGGTTCTCTCCAG 19597
DB 12922 CCGAGTGGCGTGGCGGACATCGACTGGAGCGGCTTCTTCTGGGCTACAGCA 12981
QY 19598 CGTC---GGTCTTCTGAGGACCTTCCCGAGGACAGGAACTGAGAGGCGCGCA 19654
DB 12982 GGGACGCGCTTGTATCGAGAGATACCGGAGGACGCGGATGCTGGAGTCCG 13041
QY 19655 CCGTTGAGCAGCGGAGAG-----CACAAACTCCCGGCACTCTCTCATGGGTC 19702
DB 13042 GCCCGCGGACCTCGAGCGGACCGCTGCGGAACCGGAGCTTCCCGTGGCTCT 13101
QY 19703 GGTACGTTCCGAGCAGGAGAGTGTCTCAGCTGCTCGCATCCACTCCCGGCG 19762
DB 13102 TCACCGCGGTTCAGCAGGAAAGCTTCTTGGTGTGAGTCCGGAACGAGCTTCA 13161
QY 19763 TGCTCGGCGGACGACTCCGAGGCGATCCCGCGCGCTGGCTGTTCAAGGATCTAG 19822
DB 13162 TCCTCGGACATTCGCGCGCGGCGGCTGCTCCGGAACGAGCTTCAAGGATCT 13221
QY 19823 TCGACTCGCTTGGCGGCTGAGGCTTTCGAACCACTTCGACGACGAGGAGCTG 19882

Db 9199 TGCTGTGGAGACGTCTCTGGAGGCGCTCTGAACGCGCGGCGTGAAGCCCGGAGTCTGC 9258
QY 15554 GTGGCAACCAACCGCGGTCTTCCGCGGCTGACCTACCAACGACTACGCGCGCGCTTTC 15613
Db 9259 GCGGACGACACGCGCGTGTATATCGAGCCACAGCGGAGGACTAGGCGCCCGGACTGC 9318
QY 15614 CCAAGCTCCGCGAGGGTTCAGAGGGTATCTCGGGCACGGAACGCGAGGATGATCGCCT 15673
Db 9319 ACGAGTCGACGACGACTCGGGCGGCTACGTCCTGACCGGCAATACCGCGAGCGTGGCCT 9378
QY 15674 CGGGTGTGTGCTTACGCTCTCGGCTTGGAAGGTTCGGCCCTCACAGTCGACACTGCCT 15733
Db 9379 CCGCGCGCATCGGCTTACTCCCTCGGTCTGAGGGGCTCGGGTCACGGTGGACACGGCGT 9438
QY 15734 GCTCTTCGTCCTCGGTCTGCTGCACTCGGCTGTGAGGGCTCGGCTCGGGGAGTGT 15793
Db 9439 GTTCGTCTGCTGTGGGACTGCACTTGGCGGTGAGGGCTTCGCGCTGGCGAGTGTCT 9498
QY 15794 CCAATGCGCTTCGCGGGTGGCGTCACTGATGTCAACCCCGCGGGTTCGTGGAGTTT 15853
Db 9499 CACTGGCATTTGGCGCGGAGCCACGATGATGCTTCGCCCGCATGTTCTGGAGTTCT 9558
QY 15854 CGCGGACGCGGCGCTTGGCGGTGGAGCGGCGGTGCAAGGGTTCCTGGCAGCGGCTGACG 15913
Db 9559 CACGGCAAGGGGCGCTCTCCGAGGAGCGCGCTGCAAGGCGTTTCGCCGAGACGGCGGACG 9618
QY 15914 GCACCGGCTGGGGTGAAGGGTTCGGAATGCTGCTGTGGAGCGGCTGTGGAGCGCGCGC 15973
Db 9619 GCACCGGCTGGGGTGAAGGGTTCGGAATGCTGCTGTGGAGCGGCTGTGGAGCGCGCGC 9678
QY 15974 GGCTCGGTCAACGAATCTCTCGGGTGTGGTGGCAGTGGCGGTCAATCAGGACGGTGCGA 16033
Db 9679 GGTTCGGTCACTCGGGTGTGGCGGTGTGGCGGGAGTGGCGTCAATCAGGATGTGCGT 9738
QY 16034 GCACCGGCTGACGCGGCGCAACGCGCGCTCCAGAGCGTGTATCGGCTGGCGCTGG 16093
Db 9739 CGAATGGGTTCAGCGGCGCGAATGCTCGGTTCGAGAGCGGGTGTATCGGTGGCGCGTGG 9798
QY 16094 CCAACCGGACCTGACCCCGCGAGCTCGATGCGGTGGAGGCGCACGCGCACCGACCA 16153
Db 9799 CTGACCGGGTCTGGTTCCTGCTGATGTGATGTGGTGGAGGCGCATGTTACGGGACCG 9858
QY 16154 CTTTGGGCGACCGGATCGAGGCGCGAGCCCTCTCCGACCTACGGAACGAGCGCGCGC 16213
Db 9859 GGTTCGGTGTATCCGATCGAGGCTCAGGGTGTGGTGGCGAGTATGGCGAGGGCGTGGC 9918
QY 16214 GCACGAACCGCTGTGGGTGGCTCGATGAAGTGAACATCGGCGACGCGGCTGGCG 16273
Db 9919 GTGGGCGTCCGGTGGTGTGGGTTCGGTGAAGTGAACATCGGTCATACGCAAGCGCGG 9978
QY 16274 CAGGTGTGGCGGGGTCAACAAGTGTGATGGCGCTGGGAATGCTGTGCTGCGCGGA 16333
Db 9979 CTGGTGTGGTGTGATGAAGATGGTGTGGCGTGGCGTGGCGGGGTGTGGTCCGAAGA 10038
QY 16334 CGTTCATGTGATGAGCGCTCGCCGATGTGACTGTGTCGCGGGGCGGGTGCAGCTGC 16393
Db 10039 CGTTCATGTGATGAGCGCTCTGCCATGTGACTGTGCTGGTGTGAGGTGAGTGG 10098
QY 16394 TGAAGGAGCGGTGCGCTGGCGCGGGAGGGGCGGCTGGCGGGGAGGAGTGTAT 16453
Db 10099 CGGTTGAGGCGGTGCGCTGGTGC---GCGGGGTGGCGGGGTGGCGGGGCTGGTGTGCT 10155
QY 16454 CGTTCGGGCTCAGCGGCAACCGCCAGTCACTCTCAAGAGCAAGCACCGCCCAACA 16513
Db 10156 CGTTCGGGATCAGTGGCAAGATGCGCATGTATCGTGGAGGAGGCGCTG----- 10206
QY 16514 TCCGCTCAGACACACCGCGCGACGACCGCCCGGAGAGCAGCGCGCGCATGTTCCGG 16573
Db 10207 ----- 10206
QY 16574 GGAAGCGGCGGCGACGACGCGGTACCGCGGGGGAAGCGACTGTGCTGTGCGGCA 16633
Db 10207 -----CGGAGCGGAGCGGAGCGGAGCGG-----GGTCCGGGCTCTG 10245

QY 16634 GTCCAGGGGTGTGGCGGTGGCTGTGTCGGCCAAAGTTCGAGCGCGGCGCTTCGCGGCCAGG 16693
Db 10246 TTGTGGGTGTGTGCGGTGGGTGTCTCCGGCGGAGTCCGGGGCGTTCGCTGTGACAGG 10305
QY 16694 CCCAGGCGCTGACGCGCACCTCAACGACACCCCGGCGCTGACCTTCGCGCGAGCTCGGGT 16753
Db 10306 CGGCAACGCTTGGCTCGGACGT-----GTCCGGGTGAAGTGCGGTCTGATGTGGGT 10356
QY 16754 ACACCTCGCCACCGCGCGCGGTTCGACACCGCGGCCACCTCATCGCGCGCGAC 16813
Db 10357 GGTGTTGTGGCCAGAGTGGTTCGAGACCGGGCGGTGATGGTCGGCAGTGAAC 10416
QY 16814 GGGACACCTTCTGCAAGCACTCCAGGCACTGCGCGCAGCGCGAACCACCCCGCGCTCA 16873
Db 10417 TCGATGCCATGCGCGAGTGTGGCGCGCTTCGCTCGGGTGGGGTGTGCGCGGGGTGG 10476
QY 16874 TCCAGCAGCGCGCGCGCGGAGCCGGGAGCGCGGAGCGGAGAAAGACGCGATTCA 16933
Db 10477 TGTGGGTGTGGCTCGGCTGAGGGTTCGTGTGGT-----TTCC 10518
QY 16934 TCTGTCCGACAGGCGACCCCAACCGCGCGCATGSCCAACCGGCTCTACCAACCCACC 16993
Db 10519 TCTTCTGTGTCAGGGTTCGCAAGTGGGTGGGATGGCGCTGGTTCGTGGATGCGTGC 10578
QY 16994 CGGTCTTCGCGCGCGCACTCAACAGCATCTGCACCCACCTCGACCCCGCACCTCGAC 17053
Db 10579 CGGTTCGCGAGGCGGTGGCGGAGTGGCTGCGGTGTGACCCGCTTGACCGGTGGT 10638
QY 17054 CCTCTCTCCCTCTCTACCCAGGACCCCAACACCGAGACACCAACCCCTCGAAGAG 17113
Db 10639 CGCTGTGAGGTGT-----GCGCGGTGGT 10665
QY 17114 CGGCGCACTGTCTCAGCAGACCCGTAGCCCGCGCGCTCTTCGCGCTTCCAGGTGC 17173
Db 10666 GTGAGGTGTCTTGGCGGGGTGATGTTGTCAGCGCGGTGTGGCGGGTGTGGTGT 10725
QY 17174 CCTCCACCGCTCTCTACCGCGGTACACATCACCCCGCTACTACCGCGGACACT 17233
Db 10726 CACTGGCGGACCTTGGCGGTA---TTACGGTGTGGAGCTCTGCGGTGTGGGCAAT 10782
QY 17234 CCTCGGGGAATCAACCGCGCGCTGCGCGCATCTCTACCTCTACCGA---CGCA 17290
Db 10783 CGAGGCTGAGATGTCTGCGCTTGTGGCTGGGGGTGAGTCTGGCGGATGTTGCGC 10842
QY 17291 CCACCTCATCACCAACCGCGCGCTCTACCAACCATGCCCCCGCGCACCATGACCA 17350
Db 10843 GGGTGTGTGTGGAGCGCGCGCATGCGCGGATCGCTGGTGGGGCGCGCATGCT 10902
QY 17351 CCTCCACACACCCCGCACACATCACACCATCTACCGCGCACCGAAGACGCTCG 17410
Db 10903 CGGTGAGCTGCGCGCGCGCTGTCGGCACCATCTGTTGAGAGGTTCGAGCGGAGTTT 10962
QY 17411 CCATCGCGCATCAACACCCCGCTCTCTGTCATAGCGGCGACCCCGCACCGTCC 17470
Db 10963 CGTTTGGCGGTCAACGCTCGCTCTGACCGGTGTGCGGTGAGCTTCAGGCGCTCG 11022
QY 17471 AACACATCACACCTCTGCGCAACAGGATCAAAACCAAAACCTCTCCCGCAAAA 17530
Db 11023 ATGAGTGTGTGGCGGTGTGAGCGGAGGGTGTCCGGGCTCTGCTGTCTCGGTGACT 11082
QY 17531 AGCGCTTCCACTCCCGCACCAACCCCATCTCTCAACCAACTCCACACGACACCCAAA 17590
Db 11083 ATGCTCTCACTCCCGGAGATGAGACAGTACCGGATGATCTGTGGAGGCTGGCGA 11142
QY 17591 CCTCACTTACACACCCCGCACCCCGCTCATACCGCGCAACACCCCGACCGCAAC 17650
Db 11143 CGATGCTTCTACATCGCGAAGTACCGTCTTCTCTGACGCGTACGCGGAGTGGCTGG 11202
QY 17651 TCCTACCCCGCA-----CTACTGACCCCAACAGCCCGCAACACCTCTCGACTAG 17701
Db 11203 ACACACCGCTCTGATGCGGGTACTGTTTACGAATCTGCGGGAGAGCGGTCCGTTCC 11262

Qy	7090	CCCAACGGGCGCTCCCAAGCAGCGTGTGTCATCCGCGCAGGCGCTTCGCCAACCGCGCGTGTGTCG	7149
Db	8644	CCCAACGGGACCCCTCGCAGCAGCGGGTGATCCGGCAGCGCTTCGCCGACGCGCGGCTGGCA	8703
Qy	7150	GCGCGTGATGTGAGCGGTTGGAGGCCACCGGCACCGGCACCACTTTGGGCGACCCGATC	7209
Db	8704	CCCGGCGACATCGACGCGCGTTCGAGAGCGACCGGCACCGGGAACCTTCGCTGGGCGACCCCATC	8763
Qy	7210	GAGGCCAGGCGCTCTCTCGCGACTCGGACAGGACCGGTGCGCGGCGAGGCGCGCTGTGG	7269
Db	8764	GAGGCCAGGCGCTCCAGGCCACATGACGGCAAGAGAGGCGCCCGCGAAAGCGCGCTCGCC	8823
Qy	7270	CTGGGCTCGGTCAAGTCAATGTGCGTCAACACAGGCTGCGCGGCGGTTCGCGGGGTG	7329
Db	8824	ATCGGCTCCGTGAAGTCCAAATCGGACACACCGACCGCGCGCGGTGCGGCGGSCATC	8883
Qy	7330	ATCAAGATGATGGCGCTGCGGCATGTGCTGCTGCGCGGACGTTGATGTGATGAG	7389
Db	8884	ATCAAGATGTCTCGCGATGTGCCCAACCGCAACCTTCGCCAAGAACCTTCACGCCGACGAG	8943
Qy	7390	CCGTGCGCGCATGTGACTCGTCCGCGGTGCGGTGCAGCTGCTGACGGAGACGTTGCC	7449
Db	8944	CCGAGCCCGACGTCGACTGGGGAAACAGCGGCTGGCCCTGTCACCGAGCGATCGAC	9003
Qy	7450	TGGCGCGGCGGAGGGGCGGTACCGCGGGCAGGAGTGTATCATTTGCGCGTCAAGCGC	7509
Db	9004	TGGCGGCGCGCACCGGTGCG---CGCGCGCGCGCTCTCTCTTTGGGATCATAGCGGG	9060
Qy	7510	ACCAACGCCACGTCATCTCGAAGAAAGACACCGCGCGACGAGT---7554	
Db	9061	ACGAACGCGACGTCGTGCTGAGACAGCGCGCGGATGCTGCTGTGAGTGTGTTGGGGCC	9120
Qy	7555	-----CCGGGGGACACACC7569	
Db	9121	GATGAGTGCTGAGGTGCTGAGACGTAGCATGCTGGGACGCTGGAGACTCCGAG	9180
Qy	7570	GCGCGGAGGTTGACGCGGCACGACGATGAGGCTGTGCGCGCAGTCTCGGGGT---7627	
Db	9181	GTGCTGAGGGCTCTGAGGCTCCGAGGCGCCCGGCGCGCCCGGCGACCGGTGAGGCGTCC	9240
Qy	7628	-----GGCGTGGCTGTGTGCGGCAAGTCGAGCGGCGCCCTGCGCGCCG	7674
Db	9241	CTCCCGGGCACCTGCGCTGGGTGCTGTCCGCAAGGACGAGCAGTCTGTCGCGCGCCAG	9300
Qy	7675	GCCACGSCCTGCACGCGCACCTCACCGACACCCCGGC-----7713	
Db	9301	GCGCGCCCTGCACCGGTGGCTGTCCGAGCCGCGCGCCGACCTGTCCGACCGGACGGA	9360
Qy	7714	---CTGACCTCGGATGTGGAATACCCCTCGCCACGCGCGCGCGTGTTCGACAC	7770
Db	9361	CCGGCGCGCTGCGGACGTCGGGTACGCTCGCACGAGCGGTACCGCCCTTCGCGAC	9420
Qy	7771	GCGCGCACCTCATCGCGCGGACCGGGAACGTTCTCTCAAGCACTCCAGGCACTCGCC	7830
Db	9421	GCGCGCGCGGTGACCGCGCGACCGGGAAGGGTTCTTGAGCGGGCTGGCCACGCTGGCC	9480
Qy	7831	GCAAGGAGCGCCACCCCGCGTTCATCCAGCAGCGCCCGCGGGGACCGGGACCGGG	7890
Db	9481	CAGGGCGGCACCTCGGCGCACGTCCACTGTGACACCGCGCGGACCGGAC---9531	
Qy	7891	GAGGCCGAGGAAGACCGCATTCATCTGCTCGGACAGGGGACCCCAAGCGCCCGCGCATG	7950
Db	9532	-----ACCGGTTCTCTTACCGGCCAGGGCAGTCAGCGCCCGCGGCC	9576
Qy	7951	GCCCAAGGCTTACACACACCCCGCTTCGCGCGCGCGACCTCAACGACATCTGAC	8010
Db	9577	GGCGGTGAGTGTACACCGGACCCCGCTTCGCGCGCGCGCTCGACGAGATCTGCGCC	9636
Qy	8011	CACCTCGACCCCGCTCGACACACCCCTCTCCCTCCCTCCTCACCCCAAAACGACACGAC	8070
Db	9637	CACCTCGACGCTCACCTCGAAATGCCCCCTGCTCGACGTAT-----GTTTCGGGCGCG	9690
Qy	8071	AACGAGGACGCGGCACCTGCTCCAGCAGACCGGCTACGCCGACGCGCGCTCTTCGCC	8130

6427 CACCCGAAACGCTGGGCGCGCTCTGTCGACCTCCCGCCAGCCGATGCGCGCGCTTC 6486
5008 CAGACCTCACCACCAACCTCTACCCAAACCCACCAACCAACCAACCTCGCCATCGCAC 5067
6487 GCCACCTGTACCGCATCTCTCGGGGCAACCGGAGGACCGATCGCCATCGCCACC 6546
5068 AGCGGACCCACACCCCGCGCTCTACCCCAACACACCTCTACCCCAACACCAACACCC 5127
6547 ACCGACTCACGCCCGCGCTCGCGGCAACCTCCACGAGCTCGGCCACCGC 6606
5128 ACCCCACCCACCGAACCCTCTATCACCGGCGGAACCGCGCTCGCCACCCAC 5187
6607 GACTGGGACCCACCGCACCTCTCTATCACCGGCGGACCGAGCCCTCGGACCCAC 6666
5188 CTCACCCACACCTCACCAACCCACCAACCAACCTCTCTCTACACGAGCCGAAC 5247
6667 ---GCCGCACTGGATGCCCCACACCGGACCGAACAACCTCTCTCTAGCGCGAGC 6723
5248 GGGCCCCACACCCCGGACCAACACCTCTCACCCCAACTCTCAACAAAGGATCCAC 5307
6724 GCGAACAAGCCCGCGGACCAACCACTCACCGCGAATCTACCGCATCGGGCGCGC 6783
5308 CTCACCATCACACCTCGGACCAACCAACCAACCAACCTCTCTCTCAACAC 5367
6784 GTCACCATCGCGCTGCGACGCTGCGGACCCCGCCACCGCATCGCAACCTCTCTCGACGCC 6843
5368 ATCCCCCACAACACCCCTCTCACACCGTCATCACACCGGAGCATCTCTCGACGAGCGC 5427
6844 ATCCCCCGGAGACGCCCTCTACCGCGCTCTCTCACACCGCGCGGCTCTGACGAGCGC 6903
5428 ACCCTCACCAACCTCACCCCAACCACTCAACCAACCTCTCTCGGCGCAACGCCACAGC 5487
6904 ATCGTGGAACGCTGACCGCGGAGAGTCTCGGCGGCGCAACCGTCTGGAAGCGCTCGGC 6963
5488 GCCACCTCTCCACCACTCACCCCAACCAACCGGCTCTCTCTCTCTCTCTCTCTCTCTCT 5547
6964 GCTCTGGTCTCGACAGCTGACCGGAGCTCTGACCTCGAGCGTCTGCTCTCTCTCTCTCT 7023
5548 TCGCGCGCGGACCTCTCGGCGGACCGCGCAACCACTACCGCGGAGCAACGCTCTAC 5607
7024 TCGGTCTGAGCACTCTGGGCTCTCCCGGTTCAGGGGCACTACCGCCCGCAACGCTTAC 7083
5608 CTCGACGCGCTCGCCACACCGGCAACCCCACTCTCCCGGCAACGAGCATTCGCTGG 5667
7084 CTCGACGCGCTCGCGGCTCGCGCGGCGGCAACCGGCGGCTCGCGCTCTCTCTCTCTCTCT 7143
5668 GGCACCTGCAAGGAACCGACTCGTCTGATTCGGAACAGCGCGCGCATCTCTGACCGC 5727
7144 GGACCTGGGACGCTGGCGGCACTGGCGCGGCTGACCGGCTGGCGGCTGGCGCAAC 7203
5728 CGCGGTTTCGACCCATGTACCCGAGTTGGCCACCGGAGC----- 5768
7204 CACGGGTGCGCGGCACTGACCGGAACTCGCCCTGGCGGCACTGGAGTCTCGGCTCGGC 7263
5769 -----GGTACGCAAGGAGATCGGGAACCGGAGCGGATGATGCTGATTCGCGGAC 5820
7264 CGGGAACGAGACCGGATCAACGCTCGGCGGACATCGACTGGGACCGCTCTCTACCTCGGCTAC 7323
5821 ATCGACTGGGCAAGATCGAACAACCTCTCAGACGAGCACTG-----GTGAGCGG 5874
7324 TCTCTCGGTCGCGCGGAGCCCTCTGCGAGGAGCTGCGCGGATGCGGCGCATCTAC 7383
5875 GCCCGGAAGGAGGAGCGAGCTCTCAGCGCGCCCACTCTCAACCGCGG-----AGTTG 5925
7384 GCACGGGACAGCGCCAGCTCGGACAGGGGGGAGCTCTCGCCACGAGCGCGCAACCCCTG 7443
5926 CACAAACCGTGGCCCATCAGAGCTCGGCGGACCAACGGCGCGCATCTCTCGAGTCTGTA 5985
7444 GCGGAGCGCTGGCGCGCGGCTCTCGCGGAGCGTCTCGGAGAGCTCTCTCTCTCTCTCTCT 7503
5986 CGAGACCATGTGGCGGAGTCTCTCGGACAGCGGACCGGACCGGACCGGACCGGACCGGAC 6045
7504 CGGGCGAGGCGCGCGGCT 7563

6046 TCGTTCCGTGCACTCGGCTTGGATTCTACTCACGGCGCTCGAGTTCCGAAACCTGCTGATC 6105
7564 GCTTTCAGGACATCGGCTTTCGACTCGCTCGCGGTGTCGAGCTCGGCAACAGGCTGACC 7623
6106 AAGCAACAGAGACTCGGCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6165
7624 CGGCGACCGGCTCTCAGCTCGCGGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7683
6166 CTCGCGGTACACCTGCAACCACTGCGGGGCAACAGCAGCGGAGTCTG----- 6213
7684 CTCGCTGCTGCTCTCGCAGGAGTCTCTCTGCTGACGAGGAGACGCGCGGACCGCCGCGG 7743
6214 -----GCTCTCTCAGCGGAGCGGTTTACCGCGGAGGCTTCT 6249
7744 TCCGCGCGCTGCGCGGACTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7803
6250 GTCACGAGCGGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6309
7804 GACGACGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCG 7863
6310 GCGGACGACTTCTGGAATCTGATCTCTCTCGAGCAGGAGCGGATCGGCGGATTTCCCGACC 6369
7864 CCGGAGGACCTGTCGCGGATGCTGCTCGAGGGCGGAGGGCATCACGCGTCTTCCCGACC 7923
6370 GACCGCGCTGGGACCTGGAACAGCTCTACGACCCCGGACCCCGGACCCCGGACCCCGGAC 6429
7924 GACCGCGCTGGGACCTGGAACAGCTCTACGACCCCGGACCCCGGACCCCGGACCCCGGAC 7983
6430 TACACCGGAAACCGGCGGATCTCTCTACGACGAGGCGCATCTGACGCGGAAATTTCTGCGC 6489
7984 TACGTCGCGAGGGCGGCTCTCTGACGACGCGGCGGATCTGACGCGGAGTCTTCTCGGC 8043
6490 ATCAGCCCGCGGAGCCCTCGCATGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGAC 6549
8044 GTCTCGCGCGGAGCGCTGCGCATGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGAC 8103
6550 TGGGAAACCATCGAACACCGGCACTCAACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGAC 6609
8104 TGGGAGGCTTCGAGCGGCGCGGATCGAGCGCGGATCGCTGCGCGGAGCAGCAGCAGT 8163
6610 GTCTTACCGGCAACAGGACGAGTACGCACTTCGCGGTGCAACAGCGGCGGAGTCA 6669
8164 GTCTTACCGGCTCTCTCTACGAGGACTACGCGGCGCGGCTCGGCGGCGGCGGCGGCGG 8223
6670 ACCGATGTTTCGCACTGACCGGACCGCGCGGAGCTCTCTCGGTCGATCTCTGATC 6729
8224 GTGGAGGTTTACTCTGACCGGAGCAGCGCGGAGCTGCGGTCGCGGCGGATCTCGGCTAC 8283
6730 AGCTTGTGTTTGTAGGCTCTGCGGTGCTGAGACGCGGTTGTTCTCTCTGCTGTTGTTG 6789
8284 ACCTTCTGCTTCGAGGCGCGCGGACGAGCTCTGACACCGGCTGCTCTGCTGCTGCTGCTG 8343
6790 GCTTTCGATCTGCGCTGTCAGCGGTTGCTGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 6849
8344 GCTTTCGATCTGCGGCTGCGGCGCTGCGGAGCTGCGGAGCTGCGGAGCTGCGGAGCTG 8403
6850 GGTGTGAGCTGCTGCTCTCGGCTGCTCTGTTGAGTCTTCTGCGGAGCGGCGGCTG 6909
8404 GCGTGGGAGTGTATGCGGACCCCGCATGTTCTGTTGAGTCTCAGCCGTACGCGGCGGCTC 8463
6910 GCGCGGACCGGCAATTCGAAGCGTTCTCGGCGGCGGAGCGGACCGGCTGCGGCTGAG 6969
8464 GCGCGGAGCGGCGGAGCAAGGCTCTCTCGGCGGAGCGGAGCTGCGGCTGCGGCGCGGAG 8523
6970 GGTGTGGGATGCTGCTGCTGAGCGGCTCTCGGAGCGGCTCTCGGAGCGGCTGCGGAG 7029
8524 GCGTGGGCTGCTGCTGAGCGGCTCTCGGAGCGGCTCTCGGAGCGGCTCTCGGAGCGGCT 8583
7030 CTGCGGCTGCTGCTGAGGCTGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 7089
8584 CTGCGGCTGCTGCTGAGGCTGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 8643

Db 2465 TCGACTACCGCTCCACAGCCGCGAGTCTAGATCATCGAGAGCGAGCTCGCCGAGTCC 2524
Qy 725 TGCTGCCATCAACCCACAGCCCTCCGCGATCCCTTCCACTTCCGTCGTCGAGCGGGCC 784
Db 2525 TCGCCGGGTCAAGCCGCGAGTCTCGCGGTCCGCTTCTTCGACACTCGAAGCGGCT 2584
Qy 785 GCCTCGACACCCGCGAGCTAGACGCGCGGTACTGTGTACCGCAACATGTGAGCAGCGTCC 844
Db 2585 GGNATCACGAGCCGCTGCTCGAGCGCGGTACTGTGTACCGCAACCTGCGCCATCGTGTGG 2644
Qy 845 GGTTCGAGCCCGCCGCCCG---CTGCTTTCGAGAGCGGGGCCAAGAGCTTGTGTCGAGA 901
Db 2645 GCTTGGCCCGCGCTCGAGACCTTGGCCACCGACGAGGCGCTTACCCACTTCTGTCGAGG 2704
Qy 902 TGAGCCCGACCGGTCTGACCATGGGGCTTCAGAGCTCGCCCGGACCTGGG----- 956
Db 2705 TAGAGCCCAACCCGCTTCAACATGGCCCTCCCGGGACCGTACCGGTCTGTGGGACCC 2764
Qy 957 -----CGACACACCCGCA 970
Db 2765 TGGCTCGCAACAGCGGGTTCAGGACCGCTAGTCCCTCCCTCGCCGGAAGCATGGGCCA 2824
Qy 971 CGCCGACACCGGTATATGGGCGCTGCGCGCGCGCGAGGGCACCTTGGACCACTTCC 1030
Db 2825 ACGGACTCGCGGTTCGACTGAGGCGCGCTCTCCCTCCGCGACCGGCCACCACTCCGACC 2884
Qy 1031 TGACGTCTCTCGC-----CCNACTACGGGGGATGTGAGAGTCTGGCGCA 1075
Db 2885 TCCCACTACGCGGTTCAGACCGAGCGGCACATGGCTTGGCGGAGATCGAGGGCGCTCGCCC 2944
Qy 1076 CCACTCGCTCTCGGCACCGCTGACCGCGCTGTCCCGCCACGACGACGAGTCTGCTCTCC 1135
Db 2945 CGCGGGCGGCGCGGTGCGAGCCCGCTCTCCCGACGAGGGCGCGGCGCG 3004
Qy 1136 TGGACCTGTGCGCGCCACACCATGCGGTGTGAACGACGACGAAAGAGCGCACCG 1195
Db 3005 AGCTGACCGGAGCAGCAGCTCGCGGTGATCTGACAAAGTCCGGGCGCAGACGGCC 3064
Qy 1196 CGTTCGATCCGGCCCATCG-----CGAGTTTCGCCCACTCG 1234
Db 3065 AGGTGTGCGGTACCGGACAGGGCGGCGAGATCGAGGTTCGACCGGACCTTCGTGAGGCGG 3124
Qy 1235 GCTTCGACTTCGTCATGGGTGTGCAACTGCGCAACCGCTTCAGCAAGGCGCACGGGCTTC 1294
Db 3125 GTTGCACTTCTGACCGGGTGCAGCTTGGCAACCGGATCAACGCGCTTCGCGGTAC 3184
Qy 1295 GGTTCGCGGTGACGCTCATTTTCGACACACCGCGCGCGGTGCGCGCGCGCTTC 1354
Db 3185 GATGCGCGCGTCCATGATCTTCGACTTTCGCCACCCCGAGGCTCTCGCGGAGCAGCTGC 3244
Qy 1355 GGACCGCGGCTCGGCCACTTCGACGAGGACACCGGCGCGTACCGGACTCACCCAGCG 1414
Db 3245 TCTCTGTGTGACGG-----GGAGGCGCGCGCGCAACCCCGCGGTGCGAGCGGCT 3297
Qy 1415 GCCACGAGGC-ACGGCAGCGCGGACGACCCCGATCGCCATCATCGGCATGGCATGCCGT 1473
Db 3298 CCGGTGCGCGCGCGGTGCGGTGCGTGCAGACCGTGGCGATCTCGGCATGGCTGCGCG 3357
Qy 1474 TTCCCGGCGGAGTTCGGTTCCTCGAAGGACCTGTGGAGTGTGCGCGCTCGCGCGGAGAC 1533
Db 3358 CTGCGCGGTGGGTTCGCTCGCGGAGGACCTGTGCGGTGTGTGCGCGCGCGGGAC 3417
Qy 1534 GCCATCGGCGGTTCCTCCACCGACCGGATGCCACGGAACAGGCTCACGCCACGAGGAC 1593
Db 3418 GCGATCTCGAGTTCTCCGCGAGGACCGCGCTGGGACGTGAGGGGCTGTACCAACCCGGAT 3477
Qy 1594 CCCACGAGCCCGGACGTTCTATCCGACGAGGCGGTTCCTTACGACGCGCGGCGCAC 1653
Db 3478 CCGGAGACCCCGGACGTCGTACGTCGCGCACCGGCGGTTTCATCGAGAACGTCGCGCG 3537
Qy 1654 TTCGACCGCGCTTCTTCGGAATCAGTTCACGTCAGGACATCGCGATGATTCGACGACG 1713
Db 3538 TTCGACCGCGCTTCTTCGGAATCTCGCGCGGAGGCGCTTCGCTCGTTCGAGGCTCGGTTGCGCGGCTCGCG 4665

Qy 1714 CGGCTGCTGTGAGAGCTCTCGGAGGCGTTTCGAGCGGCGGGAATCGATCCGCTGTCG 1773
Db 3598 CGGCTCTCTCTCGAANCTCTCTGAGAGCGCTTCAGGAGCGCGGATCGACCCGACCTCC 3657
Qy 1774 GTACCGGGTCCCCTGACGCGGTCTTCGCGCGCGCTCTCTCTTCGACTACGCGCCCGGT 1833
Db 3658 CTGCGGGACGCGAGTCTGCGGTCTTCTACCTGGGCGATGACCCAGAGTACGCGCGGAC 3717
Qy 1834 ATGACACCGGTCTGTCGAGGCGCGCGAGCGTGGAGGGCCACATCTTCACCGGTACC 1893
Db 3718 CT-----GCGGGACGCGGGGAGGCGCTTCGACGGCTTACCTCTGACCGGCAAC 3765
Qy 1894 ACGGCGAGCTCTGTCGCGCGGTATCGCTTACAGCTTCGCGGCTTGGGAAGGCGCGCGATC 1953
Db 3766 ACGGCCAGCTGATGTGCGGCGCGGTCTCGTACACACTCGGCTTGAGGGCCCGCCCTG 3825
Qy 1954 ACCGTGACACGCGGTGTCGCGCATCTGCTGTGAGGCTGCACTCTGCGGTGCGAGTGTG 2013
Db 3826 ACGGTGACACGCGCTGCTGCTGCTGCTGCGCTGCGCTGCACTCTGCGCGTGCAGGCCCTG 3885
Qy 2014 CGGTGCGGTGAGTGCACGCTCGCGCTGCGCGCGCGCTCTCGTCACTCTCACCTCGGC 2073
Db 3886 CGCAAGGCGAGGTTCGACATGGCGCTCCCGCGCGGTGGCGGTATGCCACGCGCCGG 3945
Qy 2074 ATGTTCACTCGAGTTCTCCCGCAGCGCGGCTGTCTGCTGAGCGGAGGTGCAAGGCGTAC 2133
Db 3946 ATGTTCTGAGTTTCAGCCGCGACGCGCGGTGCGCGGAGCGCGCTCGAAGGCGTTC 4005
Qy 2134 TCGGCTGACGCGACGCGACCGGCTGCGGCGAGGCGGTGCGGATGCTGTTGTTGAGCGG 2193
Db 4006 GCGCGTTCGCGGACGCGACCACTGCTGCGAGGCGTTCGCGCTCTCTCTGTCAGGCGC 4065
Qy 2194 TTGTCGATGCGGTGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCGAGTCCGCTC 2253
Db 4066 CTGTCGAGCGCGCGCGCAACCGGACACCGAGGTCTCTGCGGTGCTCGCGCGACGCGCTT 4125
Qy 2254 AACCAAGACGCTGCTCGAATGGGCTGACGCGCGCGCGCAACCGGTTCGCGCTCAGGAGCGG 2313
Db 4126 AACCAAGACGCGCGCGAGCAACCGGCTCACGCTCCGACGCGCGCTTCGACGAGCGCTC 4185
Qy 2314 ATCCGCGAGGCTTGGCGAAGCGCGGTGCTCGTGGCGGATGTTGGATGTTGTTGAGGGG 2373
Db 4186 ATCCGCGCGCGCTGCGGACCGCGCTGACGACTCCGACGTGGAGCTCTGTCGAGGCA 4245
Qy 2374 CAGGCGAGGCGACGCTGCGTATCCGATCCGAGGACAGGCGTTCGTCGCGACGCTAC 2433
Db 4246 CAGGCGAGGCGACGCTGCGGACCGGATCGAGGCGCGGCGCTGATCCCGCTTAC 4305
Qy 2434 GGGCAG---CGGGCGGTGACAGGCGGCTGCTGGCTGGGCTCTCTGAAGTCCAACTCGGG 2490
Db 4306 GGCAGGGCGGTGACGACGACGACGCGCTGCGCTCGGCTCGTTCGAGTCCAACTCGGG 4365
Qy 2491 CACACATGCTGCTGCGCGGTGCGGTGCGGTGCTCAAGATGCTGATGCGGTGCGGGNG 2550
Db 4366 CACACCGAGCGCGCGCGCTCTCCGCTGTCATCAAGATGCTCCAGGCGATGCGGCCAC 4425
Qy 2551 GGGGTGTCGCGGAGCGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATG 2610
Db 4426 GGAATGCTGCTGCGGAGCGTTCGACGAGGCGCTTCGACGAGTTCGACGAGTTCGATG 4485
Qy 2611 GGGCGGTGCGGCTGCTGACGAGGCGGTGCGTGGCGGGGAGCGCGGAGGCGGTTG 2670
Db 4486 GGGCGGTGGAATCTCTCACGAGCGCTTCGATGCGCGGAGAGAGAGAGAGAGAGAGAG 4545
Qy 2671 CCGCGCGGCGGAGTGTCTGCTGCTGCGGATCGCGGCGACGAATGCGGATGTGATTTTGGAG 2730
Db 4546 CCGCGCGCGCGCTCTCTCTCTTCGCGATCAGCGGCGACCAATGCGCATGTGTTGCTCGAA 4605
Qy 2731 GAGGCGCGCGCGCGGGGCTGCTGTCGCGGGGTGGGCTGTTGAGAGGTCTCTCGGCT 2790
Db 4606 GAGGCGCGCGGTGTTGTTGAGGGTGTCTCGGTGCTCGAGGCGCTCGGTTGCGGCTCGCG 4665

Db 10228 GTACACGGCTGTGCGCCGCGCCGAGACGACTGTGCGACCCCGAGTACTGGGTTCGGCAC 10287
Qy 8647 GCCCGCAACACCGTGTAGCTAGCCACACACCAACCAACCTCCACCAACCGCGTCAACC 8706
Db 10288 GTCCGGGACCGCTCGCTTCTCTGACGGGTCCTGCTGCGGACCTCGGGCGTGGG 10347
Qy 8707 ACCTACATCGAACTCGGACCCGACAAACACCTTACACACCTTACACCAACCAACCTTCC 8766
Db 10348 ACCTGCTGAGCTGGGCCCCGAGCGGGTCTTACACCCCATGGCGCGACGCGCTCGG 10407
Qy 8767 AACCCCCCAC----- 8777
Db 10408 GACACCCCGGGATTCGCTGCGGCTCCCGCTGCGGTCTCCCGCGGCTTCCCGCC 10467
Qy 8778 --CACCACCTTACCTACCCACCCCGACACCAACCCCAACCAACCTTCTACCAAC 8835
Db 10468 GACTCCGCGCGCGCTCCGGCCCCGGCGCTGCTGCTGGCGCTGCTGGCGCGCAAG 10527
Qy 8836 CTGCGCAAAACCAACCACTTGGCAACCCCGACCACTACACCCA----CCACGACAAACCA 8892
Db 10528 CGTCCGAGACCGAGACCTGCGGAGCGCTTGGCGAGCGCTTGGCGAGCGGCGCA 10587
Qy 8893 CCCACACCCAC-----CCACCTCGACCTTCCCACTAC 8928
Db 10588 CCGACTGCGACGCTGTGCTGCGCGCTCCGGGCGCACCGGTGACCTTGCACGTA 10647
Qy 8929 CCTTCCAAACACCACTACTGGCTCGAAAGCACACCGGTCGCGGCAACGTTCA 8988
Db 10648 TCTTCCGGGCGACCGCTACTGCTGAGCGCCCGCGGCGCGACACCGGCTGACACC 10707
Qy 8989 GCAG----CCGGACTCGACCCCAACCAACCCCTTACTCGGCGCCACATTTGAACTGGCG 9045
Db 10708 GCGGCGCTCGGTCTCGGACCGCGACACCGCTGCTCGGCGCGTGTGACCTTCG 10767
Qy 9046 ACTGACGCTGAGCGCTTCTTGAGGCGCGTGTCTTTGAGGTGCGATCGGTGCTGCT 9105
Db 10768 GACCGGAGCGGCTGCTGCTCAACCGCGCGCTTCTTCCGTGCGACACCGCTGCTGCG 10827
Qy 9106 GACCATGCTCGCGCGACGCTGCTGCTGCGGCGCGCACCTTCTCGAACTCGCGCTT 9165
Db 10828 GACCAACCGCTCTGGGAGCGTCTGCTTCCCGCGCGCGGATGCTGAACTCGCGCG 10887
Qy 9166 CATGCGGCGACATACGTGGCTGCGACCGAGTGGATGAGCTGACGCTGATGCGCGCTG 9225
Db 10888 CAGCTGCGAGTCCCGCTGCTGCTGAGCTGCGGAGCTGACCTCTCTTGAACCGCTG 10947
Qy 9226 GTGGTCTGTGAGTGGGGTGTGAGTGTGAGGTGGGGTTCGCGCTCGGATGG---- 9281
Db 10948 GTACTGCCGAGCACGCTGGCGTCACTGCGGCTGACGCTCGGCGCGCGCGGAGAG 11007
Qy 9282 ---GGAGGGCGCGCTTGTGAGTGTGTATGCGCGGGTGGAGTGTGTGG--TGGG 9336
Db 11008 CCGGTGGGAGTTCGGCGGGGACGGCGACGGCCCGTCTCCCTCACTCGCGGCTCGCC 11067
Qy 9337 GTGGTGTGCTGGGTGGGTGTGAGCTGTCAATGCTTCCGCGGTGCTGCTGAGGCTGT 9396
Db 11068 GACGCGCCCGCGGTACCGCTGCTTCCACCGACCGGCTGCTGCTGCGCCACCGACCG 11127
Qy 9397 GCTGT 9456
Db 11128 CCGAGCTTCCGCTCGCGCGCGACCGTGGCGGCACTGTGGCGCGCGCGCGCGCGAGAG 11187
Qy 9457 GTGGATGTGATGCTGCTGACCGCTTGTGGCTGGGGCTGCTGTGTGTGTGTGTGTGT 9516
Db 11188 GTGCGCTTGAACGGTCTTACGAGCGGCTCGACGGGAAACGGCCCTCGCTTGTGTGTGT 11247
Qy 9517 TTTTCGGGCTGCGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9576
Db 11248 TTTACGGGCTGAACCGGCTGTGGCTGTGAGGTGAGGTCTTTCGCCGACATCGCGCTC 11307
Qy 9577 CC-----GAGAGGGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9624
Db 11308 CCGGCAACCAACGATCGACCGCGCGCGCGACCGCGAACCGGCGGGAGTGGCGCGCG 11367

Qy 9625 TTGCTGATGTTGTTCCAGCGGTGTGCTGCTGCTTCCGGGTGGGACGSGGTTTGG 9684
Db 11368 GCCCTTACGGATCCACCCCGCTGTGCTGACCGTTCGCTGACCGCATCGCGTTCGG 11427
Qy 9685 GAGGGGCGGGTTTCGGGAGGGTTCGGGTGCGCGGTGTGTGGGTGGGTGTGTGTGTGT 9744
Db 11428 GGTCTGCTGACGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11487
Qy 9745 CACCGGCGGGTGTGACCGGTGTGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9804
Db 11488 CACGCGCGGGTGTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11538
Qy 9805 CGTAGGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9864
Db 11539 ---GACGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11595
Qy 9865 CGTCTGAGTTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9924
Db 11596 CGGCTTACGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11642
Qy 9925 CCGCGGGGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9984
Db 11643 CGTCCGCGGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11702
Qy 9985 CAGGCGTGGGCTGCGACGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10044
Db 11703 ACAGACCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11751
Qy 10045 GTGGTGTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10104
Db 11752 -----GACGAGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11798
Qy 10105 GGTGTGGGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10164
Db 11799 -----CGGCTTCTACCCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11850
Qy 10165 CGTGTGCGGGTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10224
Db 11851 CCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11910
Qy 10225 GACGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10284
Db 11911 GGTGTACGGGCGCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11970
Qy 10285 GAGCATCGGACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10344
Db 11971 CACCTCGCGGGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12030
Qy 10345 ACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10399
Db 12031 TCCGCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12090
Qy 10400 CGTGTGTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10459
Db 12091 ACCGCGCAGACGAGAACCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12150
Qy 10460 GACTTGTGATCTCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10519
Db 12151 TGTGTACCGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12210
Qy 10520 CCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10579
Db 12211 GCTGTGACGAGCGGACCATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12268
Qy 10580 GTGCGGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10639
Db 12269 GCACCGCGCACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12328
Qy 10640 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10699
Db 12329 GCTTGGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 12388

[illegible]

Db 6115 GCGCGGTGCTCAGCGCGTGTGAGACGCGCGGCGAAGGTGAGGTGCTGAGCGCGGG 6174
Qy 4468 ACTTACCAGGAACCTGGAAACCCCTCACCCTCCCGACCAACCAACCCACCAACC 4527
Db 6175 GCGGACGAGACGCGTGGAGCCCTCGCGCGCGCTCAGCGCACTGACG 6222
Qy 4528 TGGCTCATCGGCATCCCGAAACCCAGACCCACCAACCCCAACATCCTCACC 4587
Db 6223 ----- 6222
Qy 4588 AACCTCCACCAACGAGGCATACCCCATCCCTCACCCTCAACACACACCAACAC 4647
Db 6223 ----- 6222
Qy 4648 CCCCACACCTCCACACACCTCCACACACCCGACACACCCGACACCAACACCAACC 4707
Db 6223 -----ACCGGT 6228
Qy 4708 GGAGCCATCAGCGGCTGCTCTCCCTCTCTCTCGCCCTCGAGAAACACCCACCCACCAAC 4767
Db 6229 GACGGCTTACCGCGGTGCTCTGCTCTCTG ----- 6259
Qy 4768 CCCCACACACCCACCGGACCTCTCTCAACCTCACCCTCACCACCAACCCACCAACCC 4827
Db 6260 -----ACGGACTCGTACCGAGGTGCGCTGGGTCCAGGCGCTCGGCGACGCC 6306
Qy 4828 CACCCACCAACCCCTCTGTAGCGCACCAACCGCACACCAACCCACCAACCAACGAC 4887
Db 6307 GGAATCAAGGCGCCCTGTGGTCCGTCAACCGAGGCGCGGTCTCGTCTGGAGCTCTCGAC 6366
Qy 4888 CCGCTCACACACCCCAACGAGCCCAACCTGGGACTCGCGCACCAACCTCTCTCGAA 4947
Db 6367 ACCCGCGCGACCCCGACCGGGCCATGCTCTGGGCTCGGCGGTCTGCGCTTGAG 6426
Qy 4948 CACCCACCAACCGCGCGAATCATGACCTCTCCCAACCAACCGCCACCCACCAACCTC 5007
Db 6427 CACCCGAAAGCTGGCGCGCTCTGACCTCTCGGCGCACCGGCGGCGGATCGCGCGCTC 6486
Qy 5008 CAGCACTCACCAACCTCACCACCCCAACCCACCAACCAACCAACCTCGGCATCGCAC 5067
Db 6487 GCGCACTCTGTCACCGACTCTCGGGCGCACCGGCGAGGACAGATCGGCATCGGCACC 6546
Qy 5068 ACCGCGACCAACCGCGCGCTCAGCCCGCACCAACCTCACCACCAACCAACCAACCC 5127
Db 6547 ACGGACTCAGCGCGCGCTCGCGCGCACCCCTCAGCGAGCTCGGCGCCACCG 6606
Qy 5128 ACCCGCACCCCGACCGAACCCTCATCAGCGGCGAACCGGCGCGCTCGCACCCAC 5187
Db 6607 GACTGGCAGCCCGACCGCACCGCTCATCAGCGGCGCACCGGAGCCCTCGGCGACCC 6666
Qy 5188 CTCACCCACCTCACCACCGCACCAACCCACCAACCTCTCTCAGCGCGAACC 5247
Db 6667 ---GCCGCACTGGATGGCCCAACCGAGCGGAACCTCTCTCTGTCAGCGCAGC 6723
Qy 5248 GGCGCCACACCCCGACCAACACCTCACCACCAACCTCAACAAAGGCATCCAC 5307
Db 6724 GCGACACAGCCCGGAGCCCAACCTCAGCGCGAATCACCAGCTCGGCGCGCG 6783
Qy 5308 CTCACCATCAGCACTCGGACACAGCAACCCAGACCAACTCAACAACTCTCTCAACAC 5367
Db 6784 GTCACCATCGCGCTCGGACGTGCGGACCCCGGCGGATGCGGCGGCGGCTCTCTCGACGCC 6843
Qy 5368 ATCCCGCACACACCCCTCACCACCGTCATCCACCGCAGGCATCTCGACGAGCC 5427
Db 6844 ATCCCGCGGAGAGCCCTCTACCGCGTGTCTCACACCGCGCGCGGTCTCGACGAGCG 6903
Qy 5428 ACCCTCACCAACCTCACCACCCCAACCTCAACAGGTCTCTCGCGCCAAAGCCACAGC 5487
Db 6904 ATCGTGACAGCTGACCGCGAGCAGGTCCGCGGCGCCACCGTGGAGGCGCTCGGC 6963
Qy 5488 GCCCACTCTCTCAACCACTCACCAACCAACACCCCGCTCAGCGCTCTCTCTACTCC 5547
Db 6964 GCCTCGGTGCTCAGCAGGTGACCCGCGGACCTCGACCTCGACCGCTCTCTCTCTCG 7023

Qy 5548 TCCGCGCGCGCACCTTCGGCGCACCCGCGCAAGCCAACTACGCGCAGCAAGCCCTTAC 5607
Db 7024 TCCGTGTCAGACACTCTGGGATCCCGGTGAGGGAACTACGCGCCGCAACAGCCCTAC 7083
Qy 5608 CTCGACGCGCTCGCCACACACCGCACACCACTCTCCCGCGCACAGCATGCGCTGG 5667
Db 7084 CTCGAGGCGCTCGCGGTCTCGCGCGCACCGCGCGCTCTCGGTGCGCTGG 7143
Qy 5668 GGCACCTGGCAAGGAACCGACTCGCTGATTCGGACAGGCGCGGCATATCTCGACGCG 5727
Db 7144 GACCGTGGGACGCTGGCGGATGGCGCGGTGACGCGGTGGCGGAGCGGTGCGCAAC 7203
Qy 5728 CGCGGTTTTCGACCCATGTACCCGAGTTTGCCACGCGCAGC ----- 5768
Db 7204 CACGCGTCCCGGATGGACCCGGAATCGCCCTGGCGCACTGGAGTCGCGGTCTCGC 7263
Qy 5769 -----GTTTCAGCAGGCGATCGCGACACCGACCGCGGTATGTCTCATTCGCCAC 5820
Db 7264 CGGACGAGACCGGATCACTCGCGACATCGACTGGGACCGCTTCTACCTCGCTAC 7323
Qy 5821 ATCGACTGGACAGATCGAACACACTCTCAGACACGAGCTG-----GTGAGCGG 5874
Db 7324 TCCTCGGTGCGCGAGCCCTCTGTGAGAGTGCCTGAGGTGCGGCGCATCTCGAC 7383
Qy 5875 GCGCGGAAAGGAGCGCAGCTGTCCAGCGCCCACTCCACCGCGG-----AGTTG 5925
Db 7384 GCACGGACAGGCGCACGTCGCGACAGGCGGAGCTCGGCCAGGCGCCACCCCTG 7443
Qy 5926 CACAAACGCTGGCGCATCAGACGTGCGCGACCAACCGGCGCATGTCTGAGTCTGTA 5985
Db 7444 GCGGAGCGCTGGCGCGCGCTCGCGGAGCTGCGCGAGAGTCTCTCGGTCTCGTA 7503
Qy 5986 CGAGACATGTGGCGGAGTGTCTCGGACCGGACCGGAGCCAGGAGCTCGCGCGCACCG 6045
Db 7504 CGGCGGAGCGCGCGCTGTCTCGGATGCGTTCGCGGAGGAGTCTCGCGCGCACCG 7563
Qy 6046 TCGTTCTGCACTCGGCTTCTGATTCAGCGCGCTCGAGTTCGGAACCTCTGTGATC 6105
Db 7564 GCCTTCAAGGACATCGCTTCTGACTCGCTCGCGGTGTCTGAGTCTCGACAGGTGAC 7623
Qy 6106 AAGCAACAGGACTCGCGCTTCTGTCTGCTGCTTCTGACACCGCACCGCTCGCAAA 6165
Db 7624 CGGCGGACCGGCTCAGCTGCGCGAGCTGCTCTCGACACCGCGCGCTGGCG 7683
Qy 6166 CTCGCGGTACACTTCAGAACCACTCGGCGCACAGCGGAGTCTG----- 6213
Db 7684 CTCGTGCTCGCTCTCCGACGAGTCTCTCGGTGACGAGGAGACGCGCGCGG 7743
Qy 6214 -----GCTCTTTCAGCGCGACCGTTTACCGCGGAGGCTTCT 6249
Db 7744 TCCGCGCGCTGCGCGACTGTCTGCGGTGCGCGCGCGCGCGGACCGATGCC 7803
Qy 6250 GTTACCGAGCGATCGCCATCGTTGGCATGCGCTGTCTGTTCCCGGCGGAGTCTCG 6309
Db 7804 GACGACGATCGATCGCATGCTGCGATGAGTCTCGCTACCGCGGTGACATCGCGACG 7863
Qy 6310 GCGGACGACTCTGCGATCTGATCTCTCCGAGCAGGAGCGCATCGGCGGATTCGCCAC 6369
Db 7864 CCGGAGGACCTGTGCGGATGCTTCCGAGGCGCGGAGGATCACGCGGTTCGCCAC 7923
Qy 6370 GACCGCGCTGGGACTCTGACACGCTCTACGACCCCGACCCCGACCAACCCCGCACCTGC 6429
Db 7924 GACCGCGCTGGGACTCTGACGCGCTGTACGACCGCGACCCCGGACCGCTCGGCGAGGCG 7983
Qy 6430 TACACCGAAACCGCGGATTCCTTACGACGAGGCGCATCTCGACCGCGAAATCTTCGCG 6489
Db 7984 TAGTCCCGAGGCGGCTTCTGACGACGCGGCGGAGTTCGACGCGGAGTCTTCGCG 8043
Qy 6490 ATCAGCGCGGAGGCGCTCGCCATGACCCCGAGCAACGACTCTCTCTCGAAACCGCG 6549
Db 8044 GTCTCGCGCGGAGGCGCTGCGCATGGAACCCGACGAGCGGATGCTCTCTGACGAGTCT 8103

Db 1940 GTGCCAACTGCTGGAGCTCTTCGCGGCTGTTCGCGCGGCCATGCGCGAATCGAGGCCG 1999
Qy 188 CGTTTCGCGCCCTACGCTGAGTCTGCTCGGTGAGCAGGTGTTCGGGAGTCTCGCGGAGCTC 247
Db 2000 CACTCTCCCGTACGTCTGCTGCTGAGGCCGTCTGTAAGGCCCCCGGTGCGC 2059
Qy 248 CCGGGGTGACACCGGGTGGACGTCTGTAAGCCGACCTGTTCGCGGCTCATGATCTCCCTGG 307
Db 2060 CCAAGCTGAGCGGGTGGATGCTGACAGCTGTGACGCTGTGACGCTTCGCGGCTCATGGTCTCGCTGG 2119
Qy 308 CCGCCCTCTGCGCTCGCAAGGGGTGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
Db 2120 CTCGCTGTGGCAGCACCAAGCGGGTGAACGCCCGAGCGCTGTGCGGCCATCTCGAGGGCG 2179
Qy 368 AGATCGCGCAGACCCACGCTCTCGGAGGCTGTGCTCCGCGACGCGCGCACGCGTGGTGA 427
Db 2180 AGATCGCGCGGGTACGTCTGCGGTGCTCTGAGCCTTGACGACGCGCGCTGTGTCTGTA 2239
Qy 428 CGCTTTGGAGCCAGGCACAGACAC---CTTTGCGGGGACCGCGCGCTGTCTCGTCTG 484
Db 2240 CCTGCGCAGCAAGTCCATCGCGGCCCACTCGCGGCAAGGCGGCGCATGCTGTCTCTCG 2299
Qy 485 CCGCCACGCGGATGAGCTCTCTCCCGGAATCGCTCTGCGACCGAGGACCAACCCGCGCG 544
Db 2300 CGCTGAGGAGGACCGCTCTGAGCGACTGCGCGGGTTGAGC-----2344
Qy 545 GGTCTGCGCTCGCAGCGCTCAAGGACCCCGGAGCACAGTCTGTTTCGGTTCGCGGAGG 604
Db 2345 GGTCTGCTGCGCGCTGTGACCGGCGCCACCGCCACCGTGTCTCGGTGACCCCGTAC 2404
Qy 605 CGCTCGGACCTGTGTGGCGGACCTCACCGCGCGCAGGTGCGCACGCGCATGATCCCG 664
Db 2405 AGATCGAAGAGCTTGTCTGCGGCTGTGAGCGCATGGGCTCGTGC CGGGTCATTCCCG 2464
Qy 665 TGGACGTTCCCGCCACTCCCGCTGATGTACGCCATCGAGNACGGGTCTGTCAGCGGCC 724
Db 2465 TCGACTACGCGTCCACAGCGCGCAGGTTCGAGATCATCGAGAGCGAGCTCGCGGAGTCC 2524
Qy 725 TGTGCGCCATCACCCACGCGCTCCCGCATCCCTTCACATCTCTCGGTGACCGCGGCC 784
Db 2525 TCGCGGGCTCAGCCCGCAGGCTCGCGGTGCGCTTCTCTCGACACTCGAAGGCGCT 2584
Qy 785 GCTCGACACCCCGAGCTAGACGCGGCTACTGTGTACCGCAACATGTTCGAGCACGCTCC 844
Db 2585 GGATCACCGAGCCGCTGCTGACGCGCGCTACTGTAGTACCGAACCTCGCCCATCGTGTGG 2644
Qy 845 GGTTCGAGCCCGCGCGCG---CTGCTTCGAGCAGGGGCCAAGAGCTTCGTTCGAGA 901
Db 2645 GCTTCGCGCGCGCTCGAGACCTTGGCCACGACGAGGGCTTCACCCACTTCGTTCGAG 2704
Qy 902 TGAGCCCGACCGGTGCTGACCATGGGCTTCAGGAGCTCGCGCCCGGACCTGGG----956
Db 2705 TCAGCGCCACCCGCTCTCACCATGGCTCTCCCGGACCGTACCGGTCTGGCGACCC 2764
Qy 957 -----CGACACACCGGCA 970
Db 2765 TGGCTCGGACAAAGCGGTGAGGACCGCTAGTCTGCTCCCTCGCGGAAGCATGGGCCA 2824
Qy 971 CCGCCGACACCGTATCATGGGACGCTGGCGCGCGCCAGGCGACCTTGGACCACTTCC 1030
Db 2825 ACGGACTCGCGGTGAGCTGAGAGCCGCTCTCTCCCTCCCGGACCGGCCCACTCCGAC 2884
Qy 1031 TGACGTCTCTCG-----CCAACTACGGGGGATGGTGAAGCTCGGCGA 1075
Db 2885 TCCCGACCTACGGTTCCAGACCGAGGCGCACTGGCTGGCGGAGATCGAGGCGCTCGCC 2944
Qy 1076 CCAACGCTCTCTCGGACCGCTGACCGCGCTGCTCCCGACGACGACGAGTCTGCTCTCC 1135
Db 2945 CCGCGGGCGAGCGCGCGGTGACCGCCGCTCTCCGACGAGGCGCGCGGCGCG 3004
Qy 1136 TGGACCTGTGCGCGGCCACACCATGCGGTGCTGAACGACGAGGAAAGCGGACCG 1195
Db 3005 AGCTCAGACCGGACGAGCAGCTGCGGTGATCTCTGGACAGGTTCGCGGCGCAGACGCGCC 3064

Qy 1196 CQTGAGTGC CGGGCCCATCG-----CGAGTTTCGCCCACTCG 1234
Db 3065 AGGTGCTGGGTACGCGACAGCGGGCAGATCGAGGTCTGACCGGACCTTCGTGAGGCG 3124
Qy 1235 GCTTCGACTCGTCTGCTGAGTCTGAGTCTGCGCAACCGCTCAGCAAGGCCACGCGCTTC 1294
Db 3125 GTTGACCTCTGACCGGGTGGACCTTGCACCAACCGATCAACCGCCCTTCGTGCGGTAC 3184
Qy 1295 GTTTCGCGTGAAGCTCATCTTCGACACACACACCGCGCGCGGTGCGCGCGCTTC 1354
Db 3185 GATGCGCGCTGCTGATCTTCGACTTCCACACCCCGAGGCTCTCGCGGAGAGCTTC 3244
Qy 1355 GACCGCGCGCTCGGCCACCTTCGACGAGGACACCGCGCCCGTACCGGATCACCGAGCG 1414
Db 3245 TCTCTGCTGACCG-----GAGCGCGCGGCAACCGCGCGGTGCGAGCGCGCT 3297
Qy 1415 GCCAGGAGGC-ACGGCAGCGCGGACGACCGGATCGGCATCATCGGATGCGCATGCCGT 1473
Db 3298 CCGGTGCGCGCGCGGTGCGCTTCGACGAGCGGTGGCGATCGTCGCGCATGCGCTTCGCGC 3357
Qy 1474 TTTCCGCGCGGAGTCCCGTCCCGAAGGACCTGTGGGAGCTGGCGCGCTTCGCGCGGAGAC 1533
Db 3358 CTGCGCGGTGGGTTCCTTCGCGGAGGACCTGTGCGCGGTGGTGGCGCGCGCGGAGAC 3417
Qy 1534 GCCATCGCGCGCTTCTCCACCGACCGCGGATGGGCCACCGAAACAGCGTACCGCCAGGAC 1593
Db 3418 GCGATCTCGAGTTCCTCCGACGACCGCGCTGGGACGTGGAGGCTGTATCACCCCGAT 3477
Qy 1594 CCCAGCAGCCCGGACGCTTATCCGAGGAGGCGCGGTTCCTTTCAGACGCGCGGCGAC 1653
Db 3478 CCGGAGCACCCCGGACGCTGCTGCGCGGAGCGGTTTCATCGAGAACGCTCGCCCGCG 3537
Qy 1654 TTTCCAGCGCGCTTCTCGGAATCAGTCCAGTGAGGACCTGGCGATGATGATCCGAGCAG 1713
Db 3538 TTTCCAGCGCGCTTCTTCGCGGATCTGCGCGCGGAGGCTTCGCCATGAGACCCGAGCAG 3597
Qy 1714 CGGCTGCTGCTGGAGACGCTCTGGAGGCGTTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
Db 3598 CGGCTCTCTTCGAAACCTCTTCGGAGGCGCTTCGAGGACCGCGGATCGACCCGACCTCC 3657
Qy 1774 GTACCGGGTTCCTGTCAGCGGCTCTTCGCGGCGCGCTCTCTTCGATACGCGCGCGGT 1833
Db 3658 CTGCGGGACGCGCAGGTTCGCGCTTCTACCTGGGCGATGACCCAGGAGTACGCGCGGAGC 3717
Qy 1834 ATGAGACCGCGTCTGCGAGGCGCGCGAGCTGGAGGCGACATCTCTACCGGTAC 1893
Db 3718 CT-----GCGGAGCGCGCGGGAAGGCTTCGACGCGCTACTGCTGACCCGCAAC 3765
Qy 1894 ACGGCGAGCGTCTCTCGGGCGGTATCGCTACAGCTTCGGGCTGGAAAGGCGCGCGATC 1953
Db 3766 ACGGCGAGCGTGTCTCGGGCGCGCTCTCGTACACACTCGGCTTCGAGGCGCGCGCTCG 3825
Qy 1954 ACCGTGAGACGCGGGTCTCGGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2013
Db 3826 ACGGTGAGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3885
Qy 2014 CCGTGGGTGAGTGAACGCTCGCGTGGCGCGGCTCTCGGTCATGTCACCTTCGCG 2073
Db 3886 CCAAGGCGAGGTGCAATGCGCTTCGCGCGCGGCTGGCGGAGCGCGCGGTTCGAAGCGCTTC 3945
Qy 2074 ATGTTTCATCGAGTTCCTCCCGCAGCGCGGCTGTCTGGTGGAGCGGAGGTGCAAGCGGTAC 2133
Db 3946 ATGTTTCGAGTTCAGCGCGGAGCGCGGCTGGCGGAGCGCGCGGTTCGAAGCGCTTC 4005
Qy 2134 TCGGCTGACCGGACGCGACCGCTGGCGGAGGCGGTGCGGATGCTGTTGTTGAGCGG 2193
Db 4006 GCGCGCTCGGCGGACGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4065
Qy 2194 TTTGCGGATGCGGTGCGGCTGCGGTCGCGGTCGCGGTCGTCGCGGTCGTCGCGGTCGCGGTC 2253
Db 4066 CTGTGCGGACCGCGCGCGCAACCGACACCGAGGTCTCTCGCGGTCTGTCGCGCGGAGCGCTTCG 4125

[illegible]

Db	12929	GCATCGCGTCTCTCGACGCGCGCCTCCGGACGACCGCCACCGGTCTGTGTCCTGC	12988
Qy	11294	TGCGTCCGCGCCGCGCCCTGCGG	11316
Db	12989	GGCTCGACGCCCGCGGCTTGGCGACGCGCGCGGAAAGAACCGGCCGGAATCCCGGGC	13048
Qy	11317	-----CCCCCTGTCAGAGACTCTCTGCCCGCACCGCGCGCGCACACCCG	11363
Db	13049	TCTTCCGGGAAGTCTGTGCGCGCAGGACGTTCGGGCCGCGCTCGCGGCTCCGCT	13108
Qy	11364	CACACACACTACCGGTGTGGCGACAAAGCGGCCCA	11399
Db	13109	CGACACACGCGGGACGGCGGCACCGCGGGACGGCGGACGGCGCGGAAACGGCGG	13168
Qy	11400	-----GCTGCACGCCCGGCTGGCGGCGAGACACAGAAACACAGCACACCACTCC	11452
Db	13169	CGGTACAGCTCGCAGACGGGCGGCACCGTGGACGGGCCGCGACGGCAGCGCTGTGC	13288
Qy	11453	TGCGCTGTGTCGCTCCACATCGCACCGTCTGGGCCACACCAACCCCGACACCATCC	11512
Db	13229	TGAGTTCTGTCGTGGCGAGGTTCGCGAAGTACTCGGCCACGCCCGGGTACCGGATCG	13288
Qy	11513	CCCCGACGGCGGTTCCGGACCTCGGCTTTCGACTTCCTTCACGCGCGTGGAACTAGCA	11572
Db	13289	ACGCCGAAACGGGGCTTCCTCGACTTCGGCTTCGACTCCCTGACCGCGGTGGAATCCGCA	13348
Qy	11573	ACCGGCTCTCCCGCACCAACCGACTCCGCTTCCCAACCACTTCCTTCGACCAACCCCA	11632
Db	13349	ACCGGCTCAATCCGCGGTGGCTTCGCCCTCCCGCGACCTTGTCTTCGACCAACCA	13408
Qy	11633	ACCCACCAACCTCACCAACCACTCCACACAACTCCAGC	11674
Db	13409	GCCGGCGGCACTTCGCTCCCACTTCGAGCGCGAGCTGCGC	13450

RESULT 8

```

US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

```

	Query Match	7.6%;	Score 2342;	DB 9;	Length 13842;
	Best Local Similarity	54.3%;	Pred. No. 0;		
	Matches 6606;	Conservative 0;	Mismatches 4530;	Indels 1026;	Gaps 52;
Qy	8	GGATGGACGGCGGGGAGAAACCCCGCCCTCGCGCAGGGGAGGTCTCGGAGTGGCCCGACG	67		
Db	1820	GGCCGGACGATCTGGCGGACGCTGGCCGGCGCTGAGGGTCTGGTCCGGGGCGTGCCTT	1879		
Qy	68	AGGCGGACGGCGGCGTCTGCTGTTTTCCCGGCGAGGGCCCGCAATGGCCGGGCATGG	127		
Db	1880	CCGGTGTCGGGCGAGTGGCGCTTCGTGTTCCCGGGGACGGCAGCGCAGTGGCGCCGATGG	1939		
Qy	128	GAAGGGGAACCTTCTCGACGCTTCGCAAGTCTTCGCGGAGAGCGTCCGCGCTTCGGAAGCCG	187		

9691 GGCAGCGCGAGCGCGCTGCTCGACGAGACGCGGTACACGAGTGCCTGTTGCGC 9750
8131 TTCCAGGTGCGCTTCCACCGCTCTCTCAGGAGGCTACCATACACCCCTACTACTAC 8190
9751 CTGAGGTGCGCTTCTCGGCTCGT---CGAGAGCTGGGTCATGCGCGCGCGACTG 9807
8191 GCGGACACTTCTCTGCGGAATACCGCGCGCTCCTCGCGGATCTCACCCTCAC 8250
9808 CTCGGTCACTCGGTGCGGAGATCGCGCGCGAGCTCGCGGTGTTCTCGCTCGC 9867
8251 GAGCCACACCTCTATACCAACCGCGCACCTCTATGCAAAACATGCCCCC---GGC 8307
9868 GAGCGCGCGCTGCTGTCGCGCGCGCGCTCATGACGAGTGGCGCGCTGCG 9927
8308 ACCATGACACCTCTCACACACCCCGCGCGCTCATGACACACCTCATCGCCACGAA 8367
9928 GGATGCTCGCGTCAGGCGCGGAGGAGAGATCCGCGTGTGGCTGGAGACGAGGAG 9987
8368 AACGAC-----CTCGCCATGCGCGCATCAACACCCCGCTCTCTGTCATCAGC 8418
9988 CGGTAGCGGGAGCTCTGAGAGCTGCGCGCGTCAACGCGCGCGAGCGCGCTCTGTC 10047
8419 GGCACCCCGCACCGTCCAAACATCACCACTCTGCGCAACAAAGGATCAAAACC 8478
10048 GCGGACGCGAGCGCGCGGAGCGAGGCTACTGTCGCGGCTCGCGCGCGAGACC 10107
8479 AAAACCTCCCGACCAACGCTTCACTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 8538
10108 CGCGGCTCGGGTACGACGCTTCCACTCCGCGCATGGAAGGATGCTCGACGGG 10167
8539 CTCACGACGACACCCAAACCTTCACTTACACACCCCGCGCGCGCTCATCGCGC 8598
10168 TTCCGCGCGCTCTGAGAGCGTGGAGTTCCGCGCGCTCTCTGACCGTGTCTCGAC 10227
8599 AACA-----CCCCACCGACCACTCTTCACTCCCGCGCGCGCGCGCGCGCGCG 8646
10228 GTACCGCGCTGCGCGCGCGCGCGAGCTGTGCGACCGCGAGTACTGGGTCCGCGAC 10287
8647 GCGCGAACACGCTGACCTAGCGACACACCGCGCGCGCGCGCGCGCGCGCGCGCG 8706
10288 GTCCGCGGACCGTCTCGCTTCTGACGCGGTCTGCTGCGGACCTCGCGGTGCGG 10347
8707 ACTTACATCACTCGGACCGCGACCAACCTTCACTCCCGCGCGCGCGCGCGCGCG 8766
10348 ACTGCTGAGTGGCGCGCGCGAGGCTCTTCACTCGCGCGCGCGCGCGCGCGCGCG 10407
8767 AACCCCGCGC----- 8777
10408 GACACCCCGCGGATTCGCTGCGCGCTCCCGCTCGGCTCTCCCGCGCGCTCTCCCGC 10467
8778 --CACCACCTTCACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8835
10468 GACTCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10527
8836 CTGCGCAAAACCACTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8892
10528 CGGTGCGGACCGGAGACCGTGGCGGAGCGCTCGCGGAGCGCGCGCGCGCGCGCG 10587
8893 CCCCACACCGAC-----CCACTCGACCTTCCCGCGCGCGCGCGCGCGCGCGCG 8928
10588 CCGACTGCGACCGCTGTTGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10647
8929 CCTTCCAAACACCACTTACTGCTCGAAGCACACCGCGCGCGCGCGCGCGCGCGCG 8988
10648 TCTTCCGCGCGCGCGCGCTTACTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 10707
8989 GCAG---CGGAGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9045
10708 GCGCGCGCTCGGTCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10767
9046 ACTGAGGTGGAGCGCTTCTGACGCGCGCTTCTTTGAGGTGCGATCCGTTGGTGGT 9105
10768 GACCGGAGCGCGCTGCTGCTCACCGCGCGCGCTCTCTCTGCGACACCGCGCGCGCG 10827

9106 GACCATGCGCTCGCGCGCACGCTGCTGCTGCGCGCGCACCTTCTCTCGAACTCGCCCT 9165
10828 GACACGCGCTCTGCGGAGCGTCTGCTCTCCCGCGCGCGCGAGTGTGAACTCGCCGG 10887
9166 CATGCGGACATACGTGGCTGCGACCGAGTGAATGAGCTGACGCTGCATGCGCGCTG 9225
10888 CACGTCGGAGTCCGCGGTCTGCGTGACGTGCGGAGCTGACCTCTTGAACCGCTG 10947
9226 GTGTTCTGTTGATGGGGTGTGAGTGTGAGTGTGGGTGGGGTTCGGCTTCGGATGG 9281
10948 GTACTGCCCGAGCACGCTGGCTCGAGCTGCGCGTACGCTGCGGGCGCGCGCGAGAG 11007
9282 --GGAGGGCGCGCTTTGGTGAAGTGTATGCGCGGGTGGAGTGTCTGTGG--TGCG 9336
11008 CCGGTGCGAGTGGCGCGGACCGCGCACCGGCCGTCTCCCTCACTTCGCGGTCTGCC 11067
9337 GGTGTCGTGCGGTGGGTGTGACGCTGTATGCTCGGGGTGCTGTTGAGGCTGCT 9396
11068 GACGCGCGCGGTACCGCTGCTGCTGCGACCGAGCCGCTGCTGCTGCGCACCGACCG 11127
9397 GCTGTCGTGTCGTGTCGATGTCGTCGCGGGGTGTCGCGCGCGCGGGTTCGGTGGCG 9456
11128 CCGAGCTTCCCGTGCGCCCGACCGTGCGCCCATGTGCGCGCGAGGCGCGCGAGAG 11187
9457 GTGATGTCGATGTCGTCGACCGTTGCTGCGGGCTGCTGTGTTGTTTGGGGCGGTG 9516
11188 GTGCGCTGAGAGTCTCTACGAGCGCTGACGCGGAAAGCGCTCGCTTCTGCTCGCTG 11247
9517 TTTTTCGGGCTGCTGTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 9576
11248 TTCCAGGGCTGAAACCGGTGTCGCGTACGAGGTGAGGTCTTCCGCGACATCGCGCTC 11307
9577 CC-----GGAGAGCGTGGGTGATGCGCGCTGCTTGGGCTGCTTGGGCTGCATCGCGC 9624
11308 CCGCGACACGAAATGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11367
9625 TTGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 9684
11368 GCGGCTTACGAGTACACCGCGCTGTCGACGCTTCTGTCGACGCTTCTGCGGTGCGC 11427
9685 GAGGGCGGGTTCGCGGAGGCTTTCGCGTTCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCG 9744
11428 GGTCTCGTCGACGAGCGCGCTGTCGCGCTGCTTCCACTGAGCGGTGTCACCGCTG 11487
9745 CACCGCGCGGTGTCGCTGTCGCGGTGTCGCTGTCGCTGTCGCGCGCGCGCGCGCGCG 9804
11488 CACGCGCGGTGTCGCGCGCGCGCGCGCTTCTGCTGCTGCTGCGCGCGCGCGCGCGCG 11538
9805 CGTGAGGCGGTGTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 9864
11539 ---GACGCGCTTCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 11595
9865 CGTCTGAGTTCGCGCTTGTGATATGGGTGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 9924
11596 CGGCTCACGCTGCGCGCTGTCACCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 11642
9925 CCGCGGGTTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 9984
11643 CGTCGCGGGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 11702
9985 CAGGCGTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 10044
11703 ACAGACCGCGACGCTTCTGTCGCGCGCGCGCGCGCTTCTGCGCGAG----- 11751
10045 GTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 10104
11752 -----GACGAGCTGAAGTTCGCGCGCGCTGAGTTCGCGCGGTGCGAGT 11798
10105 GGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 10164
11799 -----CGGCTCTACCCGACCTGCGCGCTGTCCTCCAGGACGTGCGCGCGCGCGCGCG 11850

; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

Query Match 7.6%; Score 2342; DB 9; Length 13842;
Best Local Similarity 54.3%; Pred. No. 0;
Matches 606; Conservative 0; Mismatches 4530; Indels 1026; Gaps 52;

QY	8	GGATGACGCGGGGGAAGAACCCCGCTCGCGCAGGGGAGTCTCGGAGTGGCCGACG	67
DB	1820	GGCGGAGCAATCTGGCGGAGCGCTGGCGCGCTGAGGGTCTGGTTCGGGGGTGGCTT	1879
QY	68	AGCGGACGCGGCGTCTGCTTTTCCGGGAGGGCCCGCAATGGCGGGCATGG	127
DB	1880	CGGTGTCGGCGAGTGGCTGCTGTTCCCGGGCAGGGCAGCAGTGGCGCGCATGG	1939
QY	128	GAAGGAATCTTCGACGTTTCGAGCTTTCGCGGAGAGCGTCCCGCGCTCGGAAGCG	187
DB	1940	GTCCGAATCTGTCGACTCTTCCGCGGTGTTCCGCGCGCCATGGCGGAATCGAGGCG	1999
QY	188	CGTTCCGCGCTACGTCGACTGCTGCTGAGCAGGTGTTGGGGACTCGCGGACGCTC	247
DB	2000	CACCTCTCCGTCAGTCTGCTGCTGAGGCGCTGCTGAGGAGGCGCCCGGTGGCG	2059
QY	248	CCGGGTGACACCGGTTGAGCTGCTCAGCGCAACCTGTTGCGCGTCATGATCTCCCTG	307
DB	2060	CCACGCTGGAGCGGGTCGATGCTGTCAGCCTGTGAGCTTCCGCGTCATGGTCTCG	2119
QY	308	CGCGCTCTGGCGCTCGCAAGGGTTCGACCGTTCGCGGTGCTGGGACACAGCGTGG	367
DB	2120	CTCGCGTGTGGCAGCACCAAGGGGTGACCGCCCGAGCGGTGCTGGCCACTCGCAG	2179
QY	368	AGATCGCGGAGCCACGCTCTCGGAGGCTGCTGCGCGACCGCGCAGCGGTGGTGA	427
DB	2180	AGATCGCGCGGTCAGTCTCGCGGTGCTGAGCTTGGACGACCGCGTCTGTCGTGA	2239
QY	428	CGCTTTGGAGCCAGGACAGAACAC---CTTGCCGGGACCGCGGGCTGCTCCGCTCG	484
DB	2240	CCCTGCGCAGCAAGTCCATCGCGCCACCTCGCGCGCAAGCGCGCATGCTGTCTCC	2299
QY	485	CGCGCAGCGGATGAGCTCTCGCGCGAATCGCTCGTGGACCGAGGACCAACCGCGC	544
DB	2300	CGTGAAGGAGGACGCGCTCTGGAGCGACTGGCGGGTTCGACG-----	2344
QY	545	GGCTCGCGTCGAGCGCTCAACGAGCCCGGAGCAGTCTGTTCCGGTGGCCCGGAGG	604
DB	2345	GGCTGTCGTCGCGCTGTGACAGGGCCACCGCCACCGGTGGTCTCGGTGACCCCGTAC	2404
QY	605	CCGTGCGGACCTGGTGGCGACCTCACCGCGCGAGGTGGCAGCGCATGATCCCGG	664
DB	2405	AGATCGAAGAGCTTGTCTGGGCGGTGTGAGGCGCATGGGGTCCGTGCGGGTCATTCC	2464
QY	665	TGAGCTTCCGCGCACTCCCGCTGATGTACGCCATCGAGGAAACGGGTGCTCAGCGG	724
DB	2465	TCGACTACGGCTCCACAGCGCGAGGTTCGAGATCATCGAGAGCGAGCTCGCGAGGT	2524

QY	725	TGTCGCCATCACCCACGCCCTCCGCACTCCCTTCCACTCTCTCGGTGACCGCGGCC	784
DB	2525	TCGCGGGCTCAGCCCGCAGGCTCGCGCGTTCCTCTCGACACTCGAAGCGGCT	2584
QY	785	GCCTCGACACCGGGAGCTAGACGGGGTACTGTACCGCAATGTCGAGCAGCGTCC	844
DB	2585	GGATCACCGAGCCGCTGCTCGACGGCGTACTGTGTACCGCAACTGCGCCATCGTGTG	2644
QY	845	GGTTCGAGCCCGCGCCCGG---CTGCTTCTGACAGAGGGCCCAAGAGTTCGTCGAGA	901
DB	2645	GCTTCGCCCCGGCGCTCGAGACCTTGCCACCGAGAGGGCTTCAACCACTTCGTCGAGG	2704
QY	902	TGAGCCCGCAGCCGCTGCTGACCATGGGCTTCAGGAGCTCGCCCGGACCTGGG----	956
DB	2705	TCAGGGCCACCCCGTCTCACCATGGCCCTCCCGGGACCGTCAACCGTCTGGGACCC	2764
QY	957	-----CGACACACCGGA	970
DB	2765	TGCGTCGGCAACCGCGCTCAGGACCGCTAGTCGCTCTCGCCGAGCATGGGCA	2824
QY	971	CGCGCAGACCGTGATATGGGACGCTCGCGCGCGGCGGACCGCTTGAGCACTTCC	1030
DB	2825	ACGGAATCGCGGTGACTGGAGCCGCTCTCCCTCCGCGACCGGCGCACCTCCGACC	2884
QY	1031	TGACGTCTCTCGC-----CCAACTACGGGGCATGCTGAGACGTCGGGA	1075
DB	2885	TCCCACTAGCGGTTCAGACCGAGCGCACTGGCTGGCGGAGATCGAGGCGCTCGCC	2944
QY	1076	CGACCGTCTCTCGGACCGCTGACCGCGCTGTTCCTCCAGCAGCAGAGTCTGCTGCTC	1135
DB	2945	CGCGGGGAGCGCGCGGTGACGCGCGCTCTCCGACGAGGCGCGCAGCGCGGG	3004
QY	1136	TGGAATGCTGGCGCGCCACACCATGGCGGTCTGAACGACGACGAGAACGAGCCACCG	1195
DB	3005	AGCTCGACCGGAGCAGCAGTGGCGGTGATCTTGGAAGTCTCGGGGCGAGCGGCC	3064
QY	1196	CGTCGGATGCGCGGCCCATCGG-----CGAGTTTTCGCCCACTCG	1234
DB	3065	AGGTCTGGGTACCGGACAGCGGGCGAGATCGAGTCCGCGGACCTTCGTGAGGCG	3124
QY	1235	GCTTCGACTCGTCAATGGGTGTGAATGCGCAACCGCTCAGCAAGGCGCAGCGGCTGC	1294
DB	3125	GTTGACCTCTCTGACCGCGGTGAGCTGCGCAACCGATCAACCGCGCTTCGCGGTAC	3184
QY	1295	GGTTCCTCGTGAAGCTCTTTCGACACACCGCGCGCGGTGCTCGCGCGCTTC	1354
DB	3185	GGATGGCGCGTCCATGATCTTCGACTTCCCGACCCCGAGGCTCTCGCGGAGCAGTGC	3244
QY	1355	GGACCGCGCGCTCGGCCACCTCGACGAGGACACCGCGCCCTACCGACTCACCCAGCG	1414
DB	3245	TCCTGCTGTGACCG-----GGAGGGCGGGCGAACCCCGCGGTGCGGAGCGGCT	3297
QY	1415	GCCACGAGGC-ACGGCAGCGCGGAGCAGACCGGATCGCGCATCATCGGATGCGGT	1473
DB	3298	CGGTGGCGGGCGCGGTGCGCTCAGCAGCGCGGTGGGATCGTCCGATGCGCTGCGCG	3357
QY	1474	TTCCCGGGCGGAGTCCGCTCCCGAGGACCTGTGGAGAGTGGCGCTCGGCGGAGAC	1533
DB	3358	CTGCCCGTGGGTTCGCTCGCGGAGGACCTGTGGCGGTGGTGGCGGGCGGAGAC	3417
QY	1534	GCCATCGGGCGGTTCGCCACACCGCGGATGGGCCACGAGAACAGCTCACGCCAGGAC	1593
DB	3418	GGATCTCGAGTTCGCCGAGGACCGCGGTGGGAGCTGGAGGGGTGTATACCCCGAT	3477
QY	1594	CCCACGAGCCCGGACGTTCTATCCGAGGAGCGGGTTCCTTTACAGACCGCGGAC	1653
DB	3478	CGGAGCACCCCGGACGCTGTAACGTCGCCAGGCGGTTTCATCGAGAACCTCGCGCG	3537
QY	1654	TTCCGAGCGCGGTCTTTCGGATCAGTCCAGTGGAGCAGCTGGCGATCGATCCGAGCAG	1713
DB	3538	TTCCGAGCGCGGTCTTTCGGATCTCGCGCGGAGGCGCTTCGATGAGACCCCGAGCAG	3597

Db 11368 GCCCTTACGGCATCACCCGCGCTTCTGACGCTTCTGTCACGCCATTCGCGGTGCGC 11427
 Qy 9685 GAGGGGCGGGTTCCGGGAGGTGTTCCGGTCCCGCTGTGTGGGTGTGTGTCGCTT 9744
 Db 11428 GGTCTGTGACAGCCCGAGCTGTCGCGGTCCCTTCCACTTGAAGGTGTACCGTG 11487
 Qy 9745 CACGGGCGGGTGTGACCGGTGTCCGGTGTGTGTCGGCTGTCCGGGCGGGCGCGG 9804
 Db 11488 CACGGGCGGTGCCGGCGGCGCGGTCTCGCTCTCGCGGGAGC----- 11538
 Qy 9805 GTGAGCGGTGTGCTGTGTGTCGGGATGAGCGGGTGTGCGGTGTGCGGTGTGTCGAT 9864
 Db 11539 ---GAGCGCTCTGCTGTCCCTGACGGACGGGAGGACGCCCGCTGTCTCCGTGGA 11595
 Qy 9865 GGTCTTGTGAGTGGGCTGTGATATGGGTGAGTGTGCTGTCTGCTGTTCGGCGGG 9924
 Db 11596 CGGCTCACGCTGCGCGCGGTACCGCGATCAGCGCGG-----GAGCG 11642
 Qy 9925 CGCGGGGTTTCGCTGTATGCGGTGACGTGGCTGAGGTGGGTCTGTGCGGTGTGTGG 9984
 Db 11643 GTGCGCGGCTGATGACCGGGTGGCTGGCGTCCGTACGCCCTCGCTCCGGCA 11702
 Qy 9985 CAGCGGTGGGTGTGACAGAGACGTGGGTGAGCGGTGTGCGGTGTGCGGGGTG 10044
 Db 11703 ACAGGACCGGACGCACTTGTACGGGGCGACCGCGCTCTCGGCAAG----- 11751
 Qy 10045 GTGGTGTGCGGTGCGCGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 10104
 Db 11752 -----GACGAGTGAAGTGTGCGCGCGCTCGGAGTTCGCGGGGTGGAAT 11798
 Qy 10105 GTGTGGGTGAGTGTGTGTGGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGGGT 10164
 Db 11799 -----CGGGCTCTACCCGACCTGGCGCGCTGTCCAGACGTGGCGCGCGGCC 11850
 Qy 10165 CGGTGTGCGGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 10224
 Db 11851 CCGGCGCGGTACCGTCTTGTGCGCGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 11910
 Qy 10225 GAGCGCGGTGAGT 10284
 Db 11911 GTGTACGGGGACGTTGTGCGCGGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 11970
 Qy 10285 GAGCATCCGACCGGTTGTCTCTCTGACCTGACACACGACCGGACCGACCTGAC 10344
 Db 11971 CACTCGCGGACCGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12030
 Qy 10345 ACCGT 10399
 Db 12031 TCGCGCGGACGATGGCGCGGAGGACCTGTGCGACGCGCGCGCTGTGTGTGTGTGT 12090
 Qy 10400 CGTGTGTGACCGGATGTGGGTGCGGTGGGAGCGGTGTGTGTGTGTGTGTGTGTGT 10459
 Db 12091 ACCGCGACACGAGAACCCCGCGCTTTCGGCTTCTGACCTGGCGACGCGCTCG 12150
 Qy 10460 GACTTGT 10519
 Db 12151 TGT 12210
 Qy 10520 CCGAGT 10579
 Db 12211 GCGCTGT 12268
 Qy 10580 GTGCGGAGT 10639
 Db 12269 GCACGCGACCGCGCTGCGCGCGGAGGACGCTCTGTGTGTGTGTGTGTGTGTGT 12328
 Qy 10640 TGT 10699
 Db 12329 GCTTGGGCGGACTGT 12388
 Qy 10700 TGT 10759

Db 12389 TGT 12448
 Qy 10760 CGTGTGGGCGGAGGT 10819
 Db 12449 CCTTGGGAGCGAGT 12508
 Qy 10820 GGT 10879
 Db 12509 CGGTACTGACGCCATCCCGCGGAACACCGCTCACCGCGTGTGTGTGTGTGTGT 12568
 Qy 10880 TGT 10939
 Db 12569 TCTTCTGCGACGCGACCTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12628
 Qy 10940 CCAAGT 10993
 Db 12629 CCAAGT 12688
 Qy 10994 CGCGGT 11053
 Db 12689 CAGCGT 12748
 Qy 11054 AGCGCGCGGCAATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11113
 Db 12749 AGCGCGCGGCAATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12808
 Qy 11114 CGGGGT 11173
 Db 12809 CGCGCTTCTCTGT 12869
 Qy 11174 CGGGACCGACCGCGGCGCATCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 11233
 Db 12869 GCCAGCGGACCTGTGCGCGGATGACGCGCGGATCGCGGATCGAGCGCGGAG 12928
 Qy 11234 CACTGT 11293
 Db 12929 GCATGT 12988
 Qy 11294 TGT 11316
 Db 12989 GGT 13048
 Qy 11317 -----CGCTGT 11363
 Db 13049 TCTTGT 13108
 Qy 11364 CACACCTACTCGGT 11399
 Db 13109 GCACGACGCGGACCGCGCGGACCGCGGAGCGCGGACCGCGGCGGAAACGCGG 13168
 Qy 11400 -----GCTGT 11452
 Db 13169 GGT 13228
 Qy 11453 TGT 11512
 Db 13229 TGT 13288
 Qy 11513 CCGCGACCGCGGT 11572
 Db 13289 ACGCCGACCGGGGT 13348
 Qy 11573 ACCGCTTCTGT 11632
 Db 13349 ACCGCTTCTGT 13408
 Qy 11633 ACCGCTTCTGT 11674
 Db 13409 GCGCGGCGGACTGT 13450

QY 7628 -----GGCGGTGGTGTGGCGCAAGTCGACGCGGCGCTTGGCGCCAG 7674
DB |||||
QY 9241 CTTCCCGGGGCACTGCGCTGGGTGTCTCGCGCAAGACGAGGATGCTGGCGGCCAG 9300
DB |||||
QY 7675 GCGCAGGCGCTGACGCGCCACCTTACCGACACCCCGGC----- 7713
DB |||||
QY 9301 GCGCGCGCTGACGCGGTGGTTCGAGCGCGCGCGCGACCTGTGCGACGCGGACGGA 9360
DB |||||
QY 7714 ---CTGACCTCGCGATGTCGATACCTTCGCCACCGCGCGCGCTGTTCGACAC 7770
DB |||||
QY 9361 CCGCGCGCTGCGGACGTCGGGTACACGCTCGCCACGAGCGGTACCGCTTTCGCGAC 9420
DB |||||
QY 7771 CCGCGCACCTCATCGCGCGGACCGGACAGCTTCTTGAAGCACTTCCAGGCACTTCGCC 7830
DB |||||
QY 9421 CCGCGCGCTGACCGCGCGGACCGGACGCGGTCTTGAAGGCTGCGGACGCTGGCC 9480
DB |||||
QY 7831 GAGGCGAGCGCCACCGCGCGTTCATCAGAGAGCGCGCGCGGCGGAGACCGGG 7890
DB |||||
QY 9481 CAGGCGGACCTCGCGCCACGCTGACCTGACACCGCGCGGACCGCAC----- 9531
DB |||||
QY 7891 GAGGCGGAGGACCGCATTTCTGCTCGGACAGGCGACCCAAAGCGCGCGCATG 7950
DB |||||
QY 9532 -----ACCGGTTCCTTTCACCGCGCGGCGAGTCAGCGCGCGCGCGCC 9576
DB |||||
QY 7951 GCGCACGGCTTACACACACCGCGCTTTCGCGCGCGCACTCAAAGCACTTCGCACC 8010
DB |||||
QY 9577 GCGCGTGAGCTGTACGACCGGACCGCGTTCGCGCGCGGCGCTCGACGAGATCTGGCC 9636
DB |||||
QY 8011 CACTCGACCGCCACCTGACACACCGCTTCTCCCGCTCTCTACCCAAAGCAACGAC 8070
DB |||||
QY 9637 CACTCGAGCGGTCACTCGAACTGCGCTGCTCGAGCTGAT-----GTTGCGCGCGGAG 9690
DB |||||
QY 8071 AAGAGGAGCGCGCGCATCTGCTCAGAGACCGCTAGCGCGCGCGCGCTTTCGCC 8130
DB |||||
QY 9691 GCGAGCGGAGCGCGCTGCTCGAGAGCGGCTACAGCGGTGCGCGCTGTTGCC 9750
DB |||||
QY 8131 TTCCAGGTGCGCTTCCACCGCTTCTCACCGCGCTACCACTACACCGCGCGCGCACTAC 8190
DB |||||
QY 9751 CTGAGGTGCGCTTCTTCGCGCTCGT---CGAGAGTGGGCGATGCGCGCGCGCACTG 9807
DB |||||
QY 8191 GCGGACACTTCTCGCGGAAATCACCGCGCGCACTTTCGCGCGCATCTCTACCTTACC 8250
DB |||||
QY 9808 CTGCGTCACTCGGTGCGGAGATCGCGCGCGCACTGCGCGGTGTGTTCTGCTCGCC 9867
DB |||||
QY 8251 GAGCGCACACCTCATACCCAAAGCGCGCGCTCATGCAACCATGCGCGCC---GGC 8307
DB |||||
QY 9868 GAGCGCGCGCTGCTGCGCGCGCGCGCGCTCATGAGAGCTGCGCGCGGTGGC 9927
DB |||||
QY 8308 ACCATGACCACTTCCACACACCGCGCGCGCGCGCATCACCGCGCGCGCGCAAC 8367
DB |||||
QY 9928 GCGATGCTCGCGCTTCCAGCGCGCGGAGGAGAGATCGCGGTGGGTGGAGCGGAG 9987
DB |||||
QY 8368 AAGGAC-----CTGCGCATCGCGCGCATCAACACCGCGCATCTCTGCTCATCAGC 8418
DB |||||
QY 9988 CGGTACGCGGAGCTGTGAGCTGCGCGCGCGTCAACGCGCGCGCGCGCTCTGTCTC 10047
DB |||||
QY 8419 GGCACCGCGCGCGTCAACACATCACACCGCTTGTGCGCAACAAAGGCACTCAAAAC 8478
DB |||||
QY 10048 GCGGACGCGGACGCGCGCGGAGGCGGCGGTACTGTGTCGGGTGCGCGCGCGGAGAC 10107
DB |||||
QY 8479 AAGACCTTCCCAACCAACCGCTTTCATCTCCCGCGCGCGCATCTCTCAACCAAC 8538
DB |||||
QY 10108 CCGCGCTGCGGTGAGCGCGCTTTCATCTCGCGCATGAGCGGATGCTCGAGCGG 10167
DB |||||
QY 8539 CTCACAGGACACCGCAACCTTACCTACCAACCGCGCGCGCGCGCGCTCTATCACCGC 8598
DB |||||
QY 10168 TTCCGCGCGCTGCGGAGCGGTGAGGTTCCGCGCGCGCTCTCTGACCGGTGCTCGAAC 10227
DB |||||
QY 8599 AACA-----CCCCACCGGACCACTCTCTACCGCGCGCGCGCGCGCGCGCGCG 8646
DB |||||
QY 10228 GTACCGCGCGCTGCG 10287
DB |||||

QY 8647 GCGCGCAACACCGTGCATACCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8706
DB |||||
QY 10288 GTCCGCGGACCGCTCGCTTCTCGACGCGCTCGTGTCTCGCGCGCGCGCGCGCGCG 10347
DB |||||
QY 8707 ACCTATACGAACTCGGACCGCGACACACCTTACCGCGCGCGCGCGCGCGCGCG 8766
DB |||||
QY 10348 ACCTGCTCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10407
DB |||||
QY 8767 AAGCG 8777
DB |||||
QY 10408 GACACCGCGCGGATTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10467
DB |||||
QY 8778 ---CACCACCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 8835
DB |||||
QY 10468 GACTCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10527
DB |||||
QY 8836 CTCGCGCAACCAACCACTTGGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8892
DB |||||
QY 10528 CGGTGCGGAGCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10587
DB |||||
QY 8893 CCGCACCG 8928
DB |||||
QY 10588 CCGGACTGCG 10647
DB |||||
QY 8929 CCGTTCACACCG 8988
DB |||||
QY 10648 TCCTTCG 10707
DB |||||
QY 8989 GCGG---CGGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9045
DB |||||
QY 10708 GCG 10767
DB |||||
QY 9046 ACTGACGCGTGGAGCGCTTCTTTCAGGCGCGCGCGCGCGCGCGCGCGCGCG 9105
DB |||||
QY 10768 GACCGGAGCG 10827
DB |||||
QY 9106 GACCATGCG 9165
DB |||||
QY 10828 GACCAACG 10887
DB |||||
QY 9166 CATGCGGCGCATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9225
DB |||||
QY 10888 CACGCTGCGGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10947
DB |||||
QY 9226 GTGCTTCTGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9281
DB |||||
QY 10948 GTACTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11007
DB |||||
QY 9282 ---GAGGCG 9336
DB |||||
QY 11008 CCGCGTGGCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11067
DB |||||
QY 9337 GGTGCTGCGTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9396
DB |||||
QY 11068 GACG 11127
DB |||||
QY 9397 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9456
DB |||||
QY 11128 CCGGAGCTTCCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11187
DB |||||
QY 9457 GTGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCG 9516
DB |||||
QY 11188 GTGCGCTGCG 11247
DB |||||
QY 9517 TTTTCG 9576
DB |||||
QY 11248 TTTCCAGGCG 11307
DB |||||
QY 9577 CC-----GAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9624
DB |||||
QY 11308 CCGCGCACCGAATGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11367
DB |||||
QY 9625 TTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9684
DB |||||

Db 17040 GCGCGGCGGAGGACGCGATCTCGGAGTTCCCCAGGACCGCGGTGGGACGTGGAGG 17099
Qy 12213 GCTCTACGACCCGAGACCCGACCGCTCAGGAACAGTTACACGCGAGCGAGGTTCT 12272
Db 17100 CTTGTACGACCCGAAACCCGAGGCGACGAGGAAAGATTACCGCGGAGCCGGATTCT 17159
Qy 12273 CTATGACGCGGCGGACTTCGACCGCGGTTCTTCGAGTTTCTACCGCGTGAAGCGTGGC 12332
Db 17160 GTACGAGCGGCGAGTTGACGCGGACTTCTTCGGGATCTCGCGCGGAGCCCTCGC 17219
Qy 12333 AATGGACCCGACGAGCGCTGCTGCTGAAACCACTTGGGAAACGTTGAAACAGGCGG 12392
Db 17220 CATGGACCCGACGAGCGTCTCTCTGGAGGCGCTCTGGAGGCGTTTCGAGCACGCCG 17279
Qy 12393 AATCGACCCGAGGTTCAGTCGCGGAGCGGACCGGAGCGGTTTCTCGTGGGATCAATCCGA 12452
Db 17280 GATCCGCGGCGCACCGCGCGGCGACCTCGTGGTGGGCTTTCACCGGCGTATGATACCA 17339
Qy 12453 GGAATACACACCGGATACACATCAGCCCTCAAGCGAGTCGAGGCTACCTGCTCAC 12512
Db 17340 CGACTACGCCACCGTCTCACCGAT---GTCCCGGAGGCGATCGAGGCTACCTGGGAC 17396
Qy 12513 TGGAGCGCGGCAAGCATTTGCTGAGCGGTATCTCTTACAACTTCGCGGCTGAAAGGCC 12572
Db 17397 CGGCAACTCCTCGGAGTGTCTGCTCGGCGCGGCTGCGGTACACGCTTGGCTGGAGGGCG 17456
Qy 12573 TGGGATCACTATCGACACCGCGTTCTCTCTGCTGCTGCGCTGCTGCTGCTGCTGCTG 12632
Db 17457 GCGCGTACCGGTGACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 17516
Qy 12633 AGCGCTCCGCTCGGTTGATGACACCATGCGCTCGAGGCGGCGCTCGCTCATGCGCCAC 12692
Db 17517 GGCCTCGGCAAGGCGAGTTCACATGCGCTCGCGCGGCGTACGCTCATGCTGAC 17576
Qy 12693 TCCCTTGTCTTACCGAGTTCTTCGCGAGCGGCGCTGCGCGCAGACGCGCGGTGCAA 12752
Db 17577 GCCCAGACCTTCTCGAGTTTACGCGCTCAGCGCGGCTGCGCGCGGCGGCTCGAA 17636
Qy 12753 GCGGTTTTCGCGCGGCGGAGCGGACCGGCTGCTGCTGCGAGGCTGGGATGCTGCTGT 12812
Db 17637 GTCCTTCTGTCACGCGCGACGCGACCGCTGCTGCTGCGAGGCGTGGCGTCTCTCTGT 17696
Qy 12813 GAGCGGCTCTCGAGCGCGCGCGCAACGCTACCGTGTCTTGGCGGCTGCTCGCGGCGAG 12872
Db 17697 CGAGCGCTGTCGACGCGCTGCAAGGCGCATCGATCTCGCGCTGCTCGGCGGAC 17756
Qy 12873 GCGCGTCAACAGGACGCGCGCAAGCAACGCGCTGACCGCACCGACCGTGTTCACAGT 12932
Db 17757 GCGCGTCAACAGGACGCGCGCGAGCAGCGCGCTTACCGGCTCGAACGCGCGTTCGAGCA 17816
Qy 12933 CAAGGTCAATCCGCGGCTTTGCGCAACGACACCTCTCTCCCTGCGGATGTCGATGCGT 12992
Db 17817 GCGCGTCAATCCGAGTGTGCGGACGCGCGGCTTACGACCTCGGCTGCGGCTGCT 17876
Qy 12993 GAGGCGCCACGCGACCGGACCACTCGGCGACCGATGCGAGGCTCAAGCGCTGTCGA 13052
Db 17877 CGAGGCGCGACGCGGCTAGCGGACTCGCGGACCGGATCGAGGCGGCGGCTATGCG 17936
Qy 13053 AGCTTACGTCAGGACCGCGCCCAACGCGCGCGCTCTGCTGCGGAACCGCTCAAGTCAA 13112
Db 17937 CAGCTACGCGGACGCGGCGTGAAGGCGAACAGCGCTGCGCTCGGCTGCTGAAGTCAA 17996
Qy 13113 CATCGGCACTCATGCGCGCTCGGCTGCGGCGGCTCATCAGATGCTGATGCGCT 13172
Db 17997 CATCGGCACTCATGCGCGCGCGGCTGCTCGGCGGCTGCTCGGCGGCTGCTCGGCGG 18056
Qy 13173 GCGGATGCTGCTGCTCGCGGACGTTGATGATGATGAGCGCTGCGCGCATGTGACTG 13232
Db 18057 GCGCAGCGGCTCTGCGGAAGCGCTTCACTGAGGAGAGCGACGCGACGAGTGGACTG 18116
Qy 13233 GTCCGCGGTGCGGCTGCTGACGAGAGCGGTGCGCTGCGCGCGG---GGGAGGG 13289

Db 18117 GTCCGCGGCGCGGTGAGCTGCTCACCGAGGCCATGAGTGTGCGCGGACAAAGGCGACGG 18176
Qy 13290 GCGGCTACGCGCGGAGGAGTGTCTATTCGCGGTACGCGCACCAAGCCGCGGCTCAT 13349
Db 18177 CGGACTGCGAGGCGCGGCTCTCTCTTTCGCGGTACGCGGACGAAACGCGACGTCGT 18236
Qy 13350 CTTGAGGAGAACCCCGCCACCAATCCCGTCAGACACACCGCGCGAGCGTCCCGG 13409
Db 18237 GCTCGAAGAGGCCCCG-----CGGCGAGG 18262
Qy 13410 AGAATCAGCGCGGAGGAGTGCCTGATGAGTGTGCGGATGAGGCTGCTGCGGAGTCCAGG 13469
Db 18263 AGACCCCTGCTCTCGAGGCGACCCCGCGCTGAGCGT-----CGGTGCGCGCGG 18314
Qy 13470 GGTGTGCGCGTGTGCTGCGCAAGTTCGAGCGCGGCGCTGCGCGCGGCGCGGCG 13529
Db 18315 CTTGTGCGGTGCTGCTGCGGAGAGACTCCGCGCGGCTGAGCGCCAGATCGGACG 18374
Qy 13530 CTTGACGCGCCACCTCACCGACACC-----CCGCGCTTCGACCTCGCGGAGTTCGGGTA 13583
Db 18375 CTTGCGCGGTTGCTCGAGGCGGTACGAGCGCGCGGATCCGCGCGGTGCTGCTG 18434
Qy 13584 GACCTGCGCGACGCGCGCGGCTGTTGACACACCGCGGCACTCTCATGCGCGCGGACCG 13643
Db 18435 CGTACTGCGCGCGGCGCGGCTGAGTTCGAGCACCGGCGGCTGCTGCTGCGACCGGACA 18494
Qy 13644 GACACCTTCTGCAAGCACTCCAGCACTCGCGCGGCGGAAACCCACCGCGCGTCTCAT 13703
Db 18495 GAGCATTTGCGGCGAGGCGTACCGCTCGGAGGACTGATACGCGGCGACCGCTC--- 18551
Qy 13704 CCACGACGCGCGCGGCGGACCGGAGCGCGGAGGCGCGAGGAAAGACCGCATTCAT 13763
Db 18552 -----GAGCTGGCGCGGTGCTGCTG 18575
Qy 13764 CTGCTCGGACGAGGCGACCAAGCGCGCGGATGCGGCGGCTCTTACCAACCGACCC 13823
Db 18576 GTTCCCGCTCAGGCGACGAGTGGCGGCGGATGCGGCGGAACTCTCGAGCTGCGAA 18635
Qy 13824 GCTCTTTCGCGCGGCACTCAAGCACTGTCACACCTCGACCGCGGCGGCTGCGACCG 13883
Db 18636 GAGTTCGCGCGGCGCATGGCGGAGTGGCGGCTCTCCGCTATGTCGATGCTG 18695
Qy 13884 CTTCTCTCCCTCTCTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13943
Db 18696 GCTGGAGCGGCTGCTCGCGGCGGCG-----GGG 18725
Qy 13944 GCGCGCATGCTGTCAGCAGACCGGCTACGCGCGCGGCGGCTTTCGCTTCCAGGTGCG 14003
Db 18726 GCGCGCGGCTGAGGCGGCTGAGCTGCTCCAGCGGCTGAGCTTCTGCTGCTCATGCTTC 18785
Qy 14004 CTTCCACCGCTCTCTACCGGCGGCTACCATCACCGCGGCGGCGGCGGCGGCGGCGG 14063
Db 18786 GCT---GGCGAAGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18842
Qy 14064 CTTGCGGCGGAAATCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14123
Db 18843 GCGGCGGAGATGCGCGCGGCTGAGTGTGCGGCGGCTTCTCTCTGAGGCGGCGGCGGCGG 18902
Qy 14124 CTTCTATCCCAAGCGGCGGCGGCTTATGCAAAACATGCGCGGCGGCGGCGGCGGCGGCGG 14180
Db 18903 GGTCTGCTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 18962
Qy 14181 CTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14240
Db 18963 CTTCTCTGCTGAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19022
Qy 14241 CATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14300
Db 19023 GATCGCGCGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19082
Qy 14301 ACATCATCAACCTCTGCGGCAACCAAGGCGATCAAAACCAAAACCTCTCCCAACCAACCA 14360
Db 19083 AGAGCTCTCAGGCGGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 19142

QY	10285	GAGCATCCGACCGGTTTGTCTCTCTCGACCTCGACACCGACACCGGACACCGACCTCGAC	10344
Db	14882	CACCTCCGGGACACCGGCTGTCTCTGTGTACCGGGGTGCGGTGCGGACCCCGAGGG	14941
QY	10345	ACCGGTGCTGTGTGTGTGGGCGC-----GTGATGTGGGCGTGTGGCGGCGGTGGTG	10399
Db	14942	TCCGGCGCGACGATGGCGCGAGGACCTGTGCGACGCGCGCGCTGTGGGTCTCGTACGG	15001
QY	10400	CGTGTGGTACCGGACGTTGGCGGTGCGTGGGAGCGGTGTGTCGCCCGACCGCTGAAAC	10459
Db	15002	ACCGGCGACACGAGAAACCCCGCGCGCTTCGCGCTTCTCGACTGGCGACACGCGCTCG	15061
QY	10460	GACTTGAATCATCCGCTGATGTTTCAGCCCGAGCGGTCCGGTGACACACAGCGCGCGGT	10519
Db	15062	TGCTACCGGACCTCGCGGTGCTCTCGAGCGGGGCTTCGCGACGAAACCGGAGCTC	15121
QY	10520	CGGACGTGCTGCCACGCGCTCCGGTGGCTGCTGCTCGGCGGTGCGTGTGATGATCGG	10579
Db	15122	GCCCTGACGACGCGACCATCAGGCTGGCGCGCTGCGCTCGCG--CCGAGACCG	15179
QY	10580	GTCCGGAGGTGTTCGCTGGTGTTCGGGTGGGTGCGTGTGAGCGGTGGGACGGGTG	10639
Db	15180	GCACCGCCGACCGCGGCTCGCCCGGAGGACGCTCTGCTGACCGCGCGACCGGCG	15239
QY	10640	TGCTGGGTGCGCGGTGGCGCGCATCTGGCTGGTGTGTGGGGTGGCGGATCTGCTGT	10699
Db	15240	GCTTGGCGGACTGTGTGCGCGGACGTGTGGCGAGTGGGGGTACACGCGCTGCTGC	15299
QY	10700	TGCTGAGCGCGGTGTCTCGGATGTCTCGGGTTCGAGGGTCTGCGGCGGAGCTGGCGG	10759
Db	15300	TGCTGAGCGCGGGGACCGGACCGCCCGCGCGCGACGAGCTCGTGACGAGCTGGAGG	15359
QY	10760	CGTTGGGGCGGAGGTGCGGATGTTGCTGTGTGATGTGGGGAGCGCGGAGGTGCTCC	10819
Db	15360	CCCTGGGACCGGACGCTCTCGGTGGCGCGGTGCGACGTCCCGACCGCGAAGCCCTCACCG	15419
QY	10820	GGCTGTGAGGGGTGTTCTGCGGGGTGTCGCTGACGGGTGTGTCATGCGGCTGGTG	10879
Db	15420	CGGTACTCGACGCCATCCCGCGCGGACACCGGCTACCGGCTGTGTCACACGCGAGGGG	15479
QY	10880	TGCTGACGATGCGACGATGCGCTCTCTCACGCCCGAGCGGTGTGGGACGCGGTTCGCGG	10939
Db	15480	TCCTCTCGACGACACCTCCCGTCCATGACGACGAGGACGTTGAAACACGTACTCGCGC	15539
QY	10940	CCAGGTGATGCGGCTCTTTGCTGATGAGCT-----GAGCGGGGTATGGAGCTGT	10993
Db	15540	CCAAGGTGACGCGCGGTTCTCTCTGACGAACTCACCTCGACGCGCCGCTACGACCTGG	15599
QY	10994	CGCGGTTGCTGTGTTCTCTCGCGCGCGGGATCCTGGGTCGCGCGGCGAGGGCACT	11053
Db	15600	CAGCGTTGCTCATGTTCTCTCGCGCGCGCGCTCTTCGTTGGCGCGGGGCGGCGCT	15659
QY	11054	AGCCCGGCGCAATGCGCTGTGACGCGCTGCGGTACCGGGCGCGGCGGGGTCTCG	11113
Db	15660	ACGCCCGCGCAACGCAACCTCGACGCGCTCGCTGGCGCGCGCGGACGCGGACTCC	15719
QY	11114	CGGGGTGTGCTGGCGTGGGGCTGTGGAGAGCGCCACCGGATGACCGGCACTCG	11173
Db	15720	CGGCCCTCTCTCTGCGTGGGGCTCTGGGCCGAGACACGCGGATGACCGGCGAGCTCG	15779
QY	11174	CGGGCACCGACACCGCGGATCATCGGTTCGGTCTGCACTCCATGTGACCCCGGACG	11233
Db	15780	GCCAGCGGACCTGCGCGGATGAGCGCGCGGGCATCGCGGGATCAGGACCGCGAGG	15839
QY	11234	CATGCGGCTCTTCGATGCGGCGCTGTGGCTGTGACCGCGGCGGTCTGCTGCCCGCGAC	11293
Db	15840	GCATCGGCTCTTCGACGCGCGCTTCGCGACGACCGCGGACCGCGGCTCTGCTGCCCTCG	15899
QY	11294	TGCGTCCCGCGCGCGCTGCGG-----GAGCGCGGAAACGACCGCGCGGAACTCCCGGCGC	11316
Db	15900	GGCTGACCGCGCGCGGCTGCGGACGCGCGCGGAAACGACCGCGCGGAACTCCCGGCGC	15959
QY	11317	-----CCCTGTCTGAGGACTCTCTGCGCGCACCGCGCGCGCGCACACCGG	11363
Db	15960	TCCTTCGGGACGTCGTTCGGCGCAGGACCGGTTCGGGCGCGCGCTTCGGCGCT	16019
QY	11364	CACACACACTACCGGTGTGCGGACAGCGC-----	11394
Db	16020	CGACGACAGCGCGGACCGCGCGCACCGCGGGAGCGGCGGACCGCGCGGAGAAACGGCGG	16079
QY	11395	--GCCAGCTGACGCGCGGCTGCGCGCGCAGACACAGAAACAGACACACACCTCC	11452
Db	16080	CGGTACGCTCGCGGACCGCGCGCGCACCGGTGAGAGGGCGCGACGCGGCGCTGCTGCG	16139
QY	11453	TCGCCCTGTGCTGCCACATCGCACCTGCGGCGCACACACACCGCGCGGCGGACCTCC	11512
Db	16140	TCGAGTTGCTGCTCGCGGAGGTTCGCGGAGTACTCGGCGACCGCGGCTCACCGGATCG	16199
QY	11513	CCCGGACCGCGGTTTCGGCGACCTCGGCTTCGACTCGCTCACCGCGGTGAACTACGCA	11572
Db	16200	ACGCGGAAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCGCTGAACTCCGCA	16259
QY	11573	ACCGGCTCTCGCGACACACCGGACTTCGCGCTTCGCCACCGCTCGGCTTCGACACCGCA	11632
Db	16260	ACCGGCTCACTTCGCGCGGCTGCGCTTCGCCCTTCGCGGACCGCTGCTTCGACACCGCA	16319
QY	11633	ACCCACACACCTCACCCACACCTTCACACAACTCC-----	11671
Db	16320	GCCCGCGGCACTGCGCTTCGCCCTTCGAGCTGCGGCGCGCTTCGGACCGG	16379
QY	11672	-----AGCCACACCGGACACCGCTGTGCG-----	11696
Db	16380	ACGAGCGGGAAACCGGAAACCGGAAACCGGAAACCGGACCGCGCTTCGGGACACCGCGG	16439
QY	11697	-----CCCGGTGTCGGGAGCTCGACAACTCGAACTCGCGCTTCGCGCTTCGACAC	11749
Db	16440	AGAAGGACCGCTGCTGCGACAACTGACCGCGCTGCGGAGGCGCTTCGCTGCGGCGG	16499
QY	11750	AAACCGACAGCGCC---AGCGAAAGAGTCACTTCGCGCTGAAGTCACTCATTTGAGG-	11805
Db	16500	TCTCGGACGCGCGCGGAGCGAAGTCTCTGGAGCACCTCGGCTCTCGCTCGATGG	16559
QY	11806	-----TGGAACGCAACCGGACGCTCGGAGCTCGGACCGCGCGGCGCGGCTGATG	11845
Db	16560	TCAGCGGCGAGACCGGAGACCGGACCGGCTCCGAGCGCGCGGCGCGGCTCGGCGG	16619
QY	11846	ACGACGA-----	11852
Db	16620	CGGAGACCGGCTTCGGCGCGCGGAGGAGCGCGGGGGGAGTGAGGACGCGCGG	16679
QY	11853	-----GAAAGTTACATCGGCAACAGAGGCTGAGATTTTCAAATTCATTGACACGACC	11905
Db	16680	GAGTCCCGGACTTCATGAACGCTTCGGCGGAGGAACTCTTCGGCGCTCTCGACGAGGACC	16739
QY	11906	TGCGGCTGTCTGAACCGGACGCTGCGCATCTTCGCGCGGTATCCG-----	11949
Db	16740	CCAGCACGCGACTGATCCCTGCGCACGCTCGCTCCCGCGGACCGCGCTCCCGGCGAC	16799
QY	11950	-----CTGGGCGCTGCTAGGACGCTGAAT	11972
Db	16800	CTCGACTCGAATCACTTCATGCGCGCTTCGGGCGCTTCAGGAACTCAAGGGGACGCGT	16859
QY	11973	GCAATGGCGAATGAAGCGAAGCTCTTCGAAATACCTCAAGCGCGTCACTCGGACCTGGA	12032
Db	16860	GTCCACGCTGAACGAAGAGAGTACCTCGACTACCTGCGTGTGCGCACCGCGGAGCTCCA	16919
QY	12033	CGGCACTCGCGTTCGCTGTACGAGGTGTGAGGCTGAGGAGGAGCGGATCGCATTTGT	12092
Db	16920	CGAGGCGGCTGCGCGCTTCGCGAGCTGAGGAGGAGGAGGAGGCGGCTGCGGATCGT	16979
QY	12093	GGGAGTGGGTGCTTACCCAGGCGGCGGAGCTCACCGCGACTGTGGCATCTCT	12152
Db	16980	CGGATGCGCTCCCGCTTCGCGCGGCGGCTTCGCGCGGAGGAGCTGTGGCGGCTGT	17039
QY	12153	CAAGTCCGAGACGCGCTATCGGGGAGTTCCCGGACCGACCGTGGATGGAACCTGGAGCA	12212

[illegible]

13799	Db	CACGCTGCGGAGTCCGCCGGTCTGCGTGAAGTGCAGCGGAGCTGACCCCTCTCTTGAAACCCGCTG	13851
9226	Qy	GTGGTTCTCTGTGGATGGGGGTGTGAGTGTGACGTTTGGGTTTGCGGTGTGCGGTGCGGATGG---	9281
13859	Db	GTACTGCCGAGCA CGGTGGCGTTCGAGCTTCCGCGTACAGGTTCGGGGCGCCGCGCGAGAG	13918
9282	Qy	---GGAGGGCGCGGTTTGGTGAAGTGTATGCGCGGGTGGAGTGTCTTGTGG--TGGG	9336
13919	Db	CCCGTGGCGAGTTCGGCCCGGGACGCGCA CGGCCCTCTCCCTCCACTCGCGCTCGCC	13978
9337	Qy	GGTGTGCGTCCGGTGGGTGTGGACGTGTATGCTTCGGGGGTGCTGGTTGAGGCTGCT	9396
13979	Db	GACGCGCCCGCCGGTACCGCCTGGTCTCTCCACGCA CGCGTCTGCTGGCCACCGACCGG	14038
9397	Qy	GCTGTGGTGTGCTGTGTGATGCTTGGCGGGGTGTGGCCGCCGGGGTGTGGCGTGTGGCTGCG	9456
14039	Db	CCGAGCTTCCCGTTCGGCCCGACCGTGTGGGCCATGTGGCCCGCCGCGAGGCGCCGAGAG	14098
9457	Qy	GTGGATGTTCGATGTGTTCGTCACCGTTTGGCTTGGGCGTGGTTGTGTTTTGGGCCCGGCTG	9516
14099	Db	GTGCCGCTCGACGGTCTCTACGAGCGGCTCGACGGGAACGGCCCTCGCTTCGTTCCGTTCCGCTG	14158
9517	Qy	TTTTTGGGCGTCCGGTGTGGGTGTGCGTGTATGGGGGGGATTTGCTGTGCTGAGGTGTGCTG	9576
14159	Db	TTCCAGGGGCTGAACCGCGGTGTGGCGGTACGAGGGTGTGAGTCTTTCGCCGACATCGCGCTC	14218
9577	Qy	CC-----GGAGGAGCGTGGGTGATGCGGCTGTGTTTGGGCTGCATCCGGCG	9624
14219	Db	CCGGCCACCAAGATGTGACCGCGCCCGCGACCGCGAAACGGCGCGGAGTGTGGCGCGCG	14278
9625	Qy	TTGCTGGATGTGTGCTCCAGCGGTTGTTCGGTGTTCCTTCGGGTGGGACGGGTTTGGG	9684
14279	Db	GCCCCCTACGGCATCCACCCCGCTCTCGACGCTTCGCTGCA CGCCATCGCGGTGCGC	14338
9685	Qy	GAGGGGGCGGGTTTCGGGAGGGTGTTCGGGTGCCGGCTGTGTGGGGTGTGTGTGCTGCTT	9744
14339	Db	GGTCTCGTCGACGAGCGCCGAGTCTGCTCCCGTCCCTTCCACTGGAACGGTGTCAACGTG	14398
9745	Qy	CACCGGGCGGGTGTACCGGTGTGGGGTGCCTGTGTGCGCTGTTCGGGCGGGCGCGCGG	9804
14399	Db	CACGCGCCGTTGCGCGCGCGCCCGGTTCCTGCTCCGCTC-----CGCGGGG	14466
9805	Qy	CCTGAGGGCGGTTCGGTCTGTGTCGGGATGAGCGGGTGTGCCGTGGCGTTCGGTCCGAT	9864
14447	Db	ACGAGCGCGTCTCGTCTCCCTGACGACGCGAGGAGCGCCGTGGTCTCGTGGAA	14506
9865	Qy	CGTCTTGATTGCGGCTGTGATATGGGTCAATTGCGTGTGCTGTCTCGTTTCGGCGGGG	9924
14507	Db	CGGCTCACGCTCGCGCCCGGTACCGCGCATCAGGCGCGGC-----GAGCCG	14553
9925	Qy	CGGCGGGTTCCGTATCCGTTGAGTGGGTGAGTGGGTCTGTGCCGTGTGTGGG	9984
14554	Db	CGTCGGCGGGTGTGATGACACGGGTGGCTTGGCGTCCGTACGCCCTCGCTTCGTCGCGCA	14613
9985	Qy	CAGGCGTGGGCGTGCAACGAGACCTGGGTGAGAGCGGTGTGGGCTGTGCCGGGGTG	10044
14614	Db	ACAGGACCGCACGCCACTTTCGTACGGGCCGACCGCGTCTCTCGGAAG-----	14662
10045	Qy	GTGTTGTTGCGGTGCCCGATGCCGGTCCCGTGTGGCGGTGTGGCGGTGTGTGCGGT	10104
14663	Db	-----GACGAGCTGAAGTCCGCCCGCCCTCGAGATCCGCGGCGCTCGAAGT	14709
10105	Qy	GGTGTGGTGAAGTTGTTGGGTGTGTTGGTGTGTTGTCAGGGGTGCTGGGGCTGGAG	10164
14710	Db	-----CGGCTCTACCCCGACTTGGCCGCGCTGTCCAGAACGTGCGCGCCGCGCC	14761
10165	Qy	CGGTTTGGCGGTTCCGCGTGTGTGTGTGACCCCGGGGTGCGGTGTGTGCGCCGCCCGGAG	10224
14762	Db	CCGGCGCCCGTACCGTACTTTCGCGCGCTGTCCCGCGGGTCCGCCGACCGCGCGCGGAG	14821
10225	Qy	GACGCCCGGTGTGATGTGGTGGGTGTGTCGTTGTGGGGCTGTGTGCTTCGCGCGCAGGCT	10284
14822	Db	GGTGTACGGGGCA CGGTGGCCCGGACGTTGAGTGTGCTCCAGGCTGTGTGGCCGACGAG	14881

Db 8675 CCGTCTCACCATGACCTTCCCGGAGACCGTCAACCGGCTCGGCACCTCTCCGTCCGAA 8734
Qy 4048 GACAAGACACCGCGCTTCTCAGCGCTTCCGACACACCCATACCAACCGCATCGC 4107
Db 8735 CAGGAGCGCAGAGCGTCTGGTCACTCTCGCGAGGCGTGGGTCAACGGGCTTCCC 8794
Qy 4108 ACACCCACCACTGGGACCACTACACCAACCAACCAACCAACCAACCAACCAACCA 4167
Db 8795 GTGGCATGGAATTCGCTCTCGCGGACCGGCTCCGCCC-----8835
Qy 4168 CACTGACCTGCGCCTACCTTCCGCTTCCAAACACGAGCACTATCTGGCTCGAGAGCTCAC 4227
Db 8836 -----CGGTCTGCCACCTACGCTTCCAGCGAGCGCTACTGGCTCGAGAACACTCC 8890
Qy 4228 CCGGGTCCCGGATTCGGGTTCGGGTGCGGTCGCGGTTCGGGTGCGGCGGGCA 4287
Db 8891 GCCG-----CCTGGCCACCG 8906
Qy 4288 GGACTCGCGGCGGAGCGGAGAGTGGAGTTCGCGGTTCGGGACCGGTGGCGCCGCGAG 4347
Db 8907 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCTCCCGCGCGGAGGGTCCGAG 8965
Qy 4348 GACTGGAACGCTCGGACCACTCTGGCGGTGCCCCCTCGCGCGCTGGACACGCTG 4407
Db 8966 CGACCGGCGCTGTCCGCGCGCTGGCTCGCGCTCAGCGCGGAGGACCACTCCGCGCAGGCC 9025
Qy 4408 GTGCCGCGACTCTCGCGCTGGACCGGACCAACCAACGACCAACCAACCAACCAACCA 4467
Db 9026 GCGCGCTGTCTACCGCGCTGGTTCGACCGCGGCGGAGGTCGAGTGTGACGCGCGGG 9085
Qy 4468 ACTTACAGAAACCTGGAACCCCTCACCTCCCGACCAACCAACCAACCAACCAACCA 4527
Db 9085 GCGGACGACGACGCTGAGGCTTCCGCGCGCGCTCACCGCACTGAG-----9133
Qy 4528 TGCTCATCGCCATCCCGAAACCCAGACCCACACCCCGGAGGTCGAGTGTGACGCGCG 4587
Db 9134 -----9133
Qy 4588 AACCTCCACCAACCGGATACCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 4647
Db 9134 -----9133
Qy 4648 CCCCACACCTCCACACACCTTCCACACACACCGGACAAAGCCCAACCAACCAACCA 4707
Db 9134 -----ACCGT 9139
Qy 4708 GAGGCAATCACCGGCTGTCT 4767
Db 9140 GACGGCTTACCGGCGTGTCT 9170
Qy 4768 CCCCACACCAACCGGACCTTCTCAACCTTCAACCTTCAACCTTCAACCTTCAACCTTCA 4827
Db 9171 -----ACGGATCTGTAACGAGGTGCGCTCGGCTCGGCGTTCGCGACGCG 9217
Qy 4828 CACCCACCAACCCCTCTGTGTAGGACCAACCAACCAACCAACCAACCAACCAACCAAC 4887
Db 9218 GGAATCAAGGCGCTTGT 9277
Qy 4888 CCGCTTCAACCCCAACCGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4947
Db 9278 ACCCGCGGACCCCGGACCGGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9337
Qy 4948 CACCCACCAACCGGCTTGT 5007
Db 9338 CACCCGGAACGCTGGGCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9397
Qy 5008 CAGCACTTCAACCAACCTTCAACCAACCGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5067
Db 9398 GCGCACTTGTACCGCACTTCTCGGCGCAACCGGCGGACCAAGATGCGCACTTGTGTGTGTGT 9457
Qy 5068 ACCGCAACCAACCGGCTTGT 5127

Db 9458 ACCGCACTCCACCGCGCGCTCTCGCGCGCACCTTCCACCGGACGTCGCGGCCACCCCG 9517
Qy 5128 ACCCCACCCCGCACGGAACACCTTATCATCACCGGGAACCGGCGCTCTCGCACCCAC 5187
Db 9518 GACTGGACCCCGACGCGCTTCTATCATCACCGGCGGACCGGAGCTCTCTCGAGCCA- 9576
Qy 5188 CTACCCACCACTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5247
Db 9577 --CGCGGACGCTGGATGGCCACCAACCGGAGCGGAACCTCTCTCTCTCTCTCTCTCT 9634
Qy 5248 GCGCCCGCACACCCCGCACGCAACACCTTCAACCAACCAACCAACCAACCAACCAACCA 5307
Db 9635 GCGCAACAGCCCGCGGACCAACCACTTCAACCGCGGACCTTCAACCGCATCGGCGCGCC 9694
Qy 5308 CTCACCATCACCACTTGGGACACGACCAACCAACCAACCAACCAACCAACCAACCAAC 5367
Db 9695 GTACATCTCGCGCTTGGAGCTCGCGGACCCCGGACCGCATCGGACCTCTCTCTCTCTCT 9754
Qy 5368 ATCCCGCCCAACACCCCTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5427
Db 9755 ATCCCGCGGAGACGCGCTTCAACCGCGTCTGTCACACCGCGCGGCTCTCGAGACGCG 9814
Qy 5428 ACCCTCAACCACTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5487
Db 9815 ATCGTGGACACGCTGACCGCGGAGGTCGCGCGGCGGCGCACCGTCCGAAGGCGCTCG 9874
Qy 5488 GCGCACTCTTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5547
Db 9875 GCTCGGTCTCGAGCTGACCGCGGACCTTCAACCGCGCTTCTCTCTCTCTCTCTCTCTCT 9934
Qy 5548 TCCGCGCGCGCACCTTCCGCGCACCCCGGCAACCAACCAACCAACCAACCAACCAACCA 5607
Db 9935 TCCGTGTGAGCACTTGGGATCCCGGTGAGGGCACTAGCGCGCGCACCAACCGCTTAC 9994
Qy 5608 CTGAGCGCTTGGCGCACACCGCGCACCAACCAACCTTCCCGCGCACCAACCAACCAAC 5667
Db 9995 CTCGACGCGCTCGCGCTCGCGCGGCGCACCGCGCGGCTCTCGGCTCTCTCTCTCTCTCT 10054
Qy 5668 GGCACCTCGGAGGAACGGACTCGCTGATTCGGACAGGCGCGCGCATATCTCTGACCGC 5727
Db 10055 GACCGTGGAGCGTGGCGGATGGCGCGCGGTGACCGGTGGCGGCGCTCTCTCTCTCTCTCT 10114
Qy 5728 CCGCGCTTTCGACCTTCAACCGGCTTGGCGACCGGCGC-----5768
Db 10115 CACGCGCTCGCGCATGACCGGAACTCGCGCTTGGCGGCTTGGAGTTCGCGCTCGCG 10174
Qy 5769 -----GTTCAACGAGCGATCGGACACCGAAACCGGCTGTATGTCTATTCGCGGAC 5820
Db 10175 CCGGACGAGACCGCGATCAACGTCGCGGACATCGACTGGGACCGCTTCTACCTCGCTAC 10234
Qy 5821 ATCGACTGGACGAGATGAAACACCTTCTCAGACGAGGACCTG-----GTGAGCGCG 5874
Db 10235 TCTCTCGTCCCGCGACCGCTTCTCGAGGAGTTCGCGGAGTTCGCGGCGGCGCATCTAC 10294
Qy 5875 CCGCGGAAAGGAGCGAGCTGTCTCAGCGCGCGCTTCAACCGGCG-----AGTTG 5925
Db 10295 GCAACGAGACGCGCATCTCGGACGAGCGGAGCTTCGCGGCGGCGGCGGCGGCGGCGGCG 10354
Qy 5926 CACAAAACGCTGGCGCATCAGACGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5985
Db 10355 GCGGAGCGCTTGGCGCGCGCGCTTCCCGGCGGAGCTGAGAGATCTCTCTCTCTCTCTCTCT 10414
Qy 5986 CGAGACATGTGGCGGAGTGTCTCGGCAACCGGAGCCGGAAGCATCTCGCGCGGCGGCGG 6045
Db 10415 CCGGCGGAGGCGCGCGCTTCTCGGATGCTTCTCGCGGAGGAGTTCGCGGCGGCGGCGG 10474
Qy 6046 TCGTCTCTGCTGCTCTGCTTCTGATTCAGTTCAGGCGGCTTCCGAAACTCTGCTGATC 6105
Db 10475 GCTTCAAGGACATCGGCTTCTGATCTGCTCGCGGTGTCTGAGTCTCGGCAACGCTGAC 10534
Qy 6106 AAGCAACAGGACTCGGCTTCT 6165
Db 10535 CCGGCGACCGGCTTCCAGCTCCCGCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10594

; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5

; LENGTH: 37948
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; us-09-988-384B-5

Query Match 9.6%; Score 2951.8; DB 9; Length 37948;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8436; Conservative 0; Mismatches 5722; Indels 1361; Gaps 69;

QY 8 GGATGACGGCGGGGAAGAACCCGCCCTCGCGCAGGGGAGGTCTCTCGAGTGGCCGACG 67
DB 4731 GGCGGACGATCTGGCGGACGCTGGCCGCGCTGAGGGTCTGGTCCGGGGGTGGCTT 4790
QY 68 AGCGGACGGCGGCGTCTTCTGTTTTTCCGGGACAGGCGCCGCAATGSCCGGGCATGG 127
DB 4791 CCGGTGTGGGGGAGTGGCGTTCGTGTTCCCGGGCAGGGCACGCAAGTGGGCGGCATGG 4850
QY 128 GAAGGAACCTCTCGACGCTTCGACGCTCTTCGGGAGAGCGTCCGGGCTCGAAGCCG 187
DB 4851 GTCCGAACTGTCTGAACTCTTCCCGCGGTGTCGGCGGCGCATGGCGAATGGGCGCG 4910
QY 188 CGTTCGCGCCTACGTCTGATGCTGCTGGAGCAGGTGTTTCGGGACCTCGCCGACGCTC 247
DB 4911 CACTCTCCCGTACGTCTGATGCTGCTGGAGCCGTCTGACGAGGCGCCCGGTGCGC 4970
QY 248 CCGGGCTGAGACCGGGTGGAGCTGCTCAGCGGACCTGTTCGCGCTCATGATCTCCCTGG 307
DB 4971 CCAAGCTGGAGCGGGTGGATGCTGTCAGCCTGTGACGTTTCGCGCTCATGCTCTGCTGG 5030
QY 308 CCGCCCTCTGGCGCTCGCAAGGGGTGAGCGGTGCGCGTGGCTGGGACACAGCCTGGGG 367
DB 5031 CTGCGGTGGGAGACACAGGGGTGAGCGCCCGAGCGGTGCTGGGCGCATCGAGGGCG 5090
QY 368 AGATCGCGGACGCCACGCTCTCGGAGGCTGTCTTCGCGCACAGCGCGCGTGGTGA 427
DB 5091 AGATCGCGCGGTAGCTGCGCGGTGCCCTGAGCTTGACGACGCGCTGCTGTGCTGA 5150
QY 428 CGCTTTGGAGCCAGGACACAGCACAC---CCTTTCGGGACCGGCGCGCTGCTCTCGTGG 484
DB 5151 CCCTGCGGACAAAGTCCATCGCGCGCCACCTCGCGGCAAGGCGGCGCATGCTGCTCCCTCG 5210
QY 485 CCGCCACCGCGATGAGCTCTCGCCCGAATCGCTCCGTGGACCGAGGACACCCGCGCG 544
DB 5211 CGCTGAGGAGGACGCGCTCTGAGCGATGGCGGGTTGACG-----5255
QY 545 GGCTCGCGCTCGCAGCGCTCAACGAGCCCGGAGCACAGTCTGTTCCGCTGCGCGGAGG 604
DB 5256 GGCTGTCCTGCGCGCTGTGAACGGGCCCAACGCCACCGTGTCTCCGCTGACCCCGTAC 5315
QY 605 CCGTTCGGACCTGTGTGGCGGACCTCAACCGCGCGAGTGGCGACGCGCATGATCCCGG 664
DB 5316 AGATCAAGAGGTGTCTCGGGGCTGTGAGGCGGATGGGGTCCGTGCGCGGGTCAATCCCG 5375
QY 665 TGGAGTTCGCCCACTCCCGCATCCCGCATCCCTTCCACTCTCTCGGTGACCGCGCGCC 784
DB 5376 TGAATACCGGCTCCCAACCGCGGAGTTCGAGATCATCGAGCAGAGCTTCGCGGAGGTCC 5435
QY 725 TGTGTGCCATCAACCCACGCGCTCCCGCATCCCTTCCACTCTCTCGGTGACCGCGCGCC 784
DB 5436 TCGCGGGCTCAGCCCGCAGGGCTCCGCGGTGCGGTCTTCTTCGACACTCGAAGGGGCT 5495
QY 785 GCCTGACACCCGCGAGCTAGACGGGCGGTACTGTGTACCGCAACATGTGAGCACGGTCC 844

DB 5496 GGATCACCGAGCCCGTGTCTCGACGGCGGCTACTGTGTACCGCAACCTGCGCCATCGTGTGG 5555
QY 845 GGTTCGAGCCCGCGCGCGG---CTGCTTCTCAGCAGGGGCCCAAGAGCTTCGTCTGAGA 901
DB 5556 GTTTCGCGCGCGCGCTGAGACCTTGGCCACCGACGAGGGCTTACCCACTTCTGTCGAGG 5615
QY 902 TGAGCCCGCACCCCGGTGTCTGACATGGGCTCTCCAGGAG-----939
DB 5616 TCAGCGCCACCCCGCTCTCACCATGGCCCTCCCGGGACCGTCAACCGGTCTGTGGGACCC 5675
QY 940 -----CTGCCCGGACCTTGGGCGGACACACCGGCA 970
DB 5676 TCGCTGCGACAAACCGGCTCAGGACCGGCTCTAGTCCCTCTCCCTCGCGAAGATGAGGCCA 5735
QY 971 CGCGCGACACCGTGTATCATGGGCGCGCTGCGCGCGCGGCGAGGCGACCTCTGAGCACTTCC 1030
DB 5736 AGCGACTCGCGGTGCACTGAGAGCCGCTCTCCCTTCGCGACCGGCGCACCACTCCGACCC 5795
QY 1031 TGACGCTCTCTCGC-----CCAACTACGGGGGCGATGTTGAGAGCTCGGCGCA 1075
DB 5796 TCCCCACCTACCGGTTCCAGACCGAGCGCCACTGGCTGGCGAGATCGAGGGGCTCGCCC 5855
QY 1076 CCACCGTCTCTTCGCGACCGCTGACCGGCTGTTCCTCCACACGAG-----1119
DB 5856 CGCGGGGCGAGCGCGCGGTGACGCGCGCTCTCCGACGAGGCGCGCGGCGCGG 5915
QY 1120 -----CAGCAGTCTGCTCTTCGACCTGCTGGCGCGCCACACCATGGGGTCTGTAACG 1174
DB 5916 AGCTCGACCGGACGAGAGCTGCGCGTATCTCTGGAAGGTTCCGGGCGGACAGCGGCC 5975
QY 1175 ACAGCGGAAACGAGCGCACCGGCTCGGATCGCGGCCCATCGCGGAGTTCGCGCCACTCG 1234
DB 5976 AGGTGCTGGGTTACGCGACAGCGCGGCGAGATCGAGGTGACCGGACCTTCCTGTAGGCG 6035
QY 1235 GCTTCGACTCTCGTCTATGGGTGTGAACTGCGCAACCGCTCTAGCAAGGCGCACGGGCTGC 1294
DB 6036 GTTGCACTCTCCCTGACCGCGGTGACCTGCGCAACCGGATCAACCGCGCTTCGCGGTAC 6095
QY 1295 GGTTCGCGCTGACGCTCATCTTCGACACACACCGCGCGCGGCTGCGCGCGGCTTC 1354
DB 6096 GGAATGGCGGTTCATGATCTTCGACTTCCCGACCCCGAGGCTCTCGCGGAGAGCTGC 6155
QY 1355 GGAACGCGCGCTCGCGCCACTCTGACGAGGACACCGCGCCCGCTACCGGACTCACCCAGCG 1414
DB 6156 TCCTGCTGTGACCG-----GAGGCGGCGCGCAACCCCGCGCTGCGAGCGGCT 6208
QY 1415 GCCACGAGGC-AAGCGAGCGGCGGACGACCGGATGCGCATATCGGCAATGCGCATGCCGT 1473
DB 6209 CCGGTGGCGGGCGCGGTGCGCTGACGAGCGGTGGCGATCTGCGCATGCGCTTCCGCGC 6268
QY 1474 TTCCCGGGGAGTCCCGTCCCGAGGACCTGTGGGAGCTGGCGCTCGGGCGGAGAC 1533
DB 6269 CTGCGCGGTGGGTGCGCTCGCGGAGGACCTGTGGCGGCTGGTGGCGCGGCGGGGAC 6328
QY 1534 GCGATTCGCGCGCTTCCCGCACCGACCGCGGATGGCCACGGAACAGCGTCAAGCGGCGGAG 1593
DB 6329 GCGATCTCGAGTTCCTCCGACGAGACCGCGCTGGGACGCTGGAGGGCTGTACACCGGAT 6388
QY 1594 CCGACGAGCCCGGACGATTTCTATCCGAGGAGGCGGGTTCCTTTCAGACGCGGCGGAC 1653
DB 6389 CCGGAGCACCCCGGACGCTCGTA GTCCTCCGAGGCGGCTTTCATCGAAGACGTCGCGCGC 6448
QY 1654 TTCCGAGCGCGCTTCTTCGGAATCAGTCCAGTGGGACCTTGGCGATGATTCGCGAGCAG 1713
DB 6449 TTCGACGCGGCTTCTTCGGAATCTCGCGCGGAGGCGCTTCGCGATGATCCCGGAGCAG 6508
QY 1714 CGGTGCTGTCTGAGACGCTCTCGGAGGCGTTTCAGAGCGGGGCGGAATCGATCTCGTCTG 1773
DB 6509 CGGTCTCTCTCGAACCCTCTTCGAGGCGCTTCGAGGACCGCGGATTCGACCCGACCTCC 6568
QY 1774 GTACGCGGCTCCCGTACGCGGCTTTCGCGGCGGCGCTCTCTCTTCGACTACGCGCGCGT 1833
DB 6569 CTGCGGGGACGCGAGGTCTGCGCTTCTCACTGGGGCGATGACCCACGAGTACGGGCGGAGC 6628

12813 GGAGCGGCTCTCCGACGCGCGCCGCAACCGTACCGTGTCTGCGCGTGTCTGCGCGCAG 12872
16527 CGAGCGCTGTCTCGACGCGCGTCCGCAAGGCGCATCGATCTCTCGCGTGTCTCGGCGCAC 16586
12873 CGCGGTCAACGAGCGCGCGCAAGCAACGCGCTGACCGGCAACCGAGCGTGTCTCAAGT 12932
16587 CGCGGTCAACGAGCGCGCGCAAGCAACGCGCTGACCGGCTGCAAGCGCGCGTTCGAGCA 16646
12933 CAAGGTATCTCCGCGCGGCTTTGCGCAACGCGTCTCTCCCTGCGCGATGTGATGCGT 12992
16647 GCGCGTATCTCGAGCTGCTCTGCGGACGCGCGCTCAAGCTCTCGAGCTCGATGCT 16706
12993 GAGGCG 13052
16707 CGAGGCG 16766
13053 AGCTACGCTACGAGCG 13112
16767 CAGGTACG 16826
13113 CATCGGCGCTCATGCGCGCTGCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13172
16827 CATCGAGACACCGAGCG 16886
13173 GCGGAATGCTCTGCTCG 13232
16887 GCG 16946
13233 GTCCGCGGTGCGGTGAGTGTGAGCGGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 13289
16947 GTCCG 17006
13290 GCGGCTACG 13349
17007 CGAGCTCG 17066
13350 CTTCGAGGAGACCG 13409
17067 GCTCGAAGAGCG 17092
13410 AGAATACG 13469
17093 AGACCGCTGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17144
13470 GGTGTGCGCGT 13529
17145 CTGTGTGCGCGT 17204
13530 CTGTGACG 13583
17205 CTCTCGCGCGTTCGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17264
13584 CACCTCG 13643
17265 CGTACTGCG 17324
13644 CGACACTTCTCGACG 13703
17325 GAGAGATTTTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17381
13704 CCACAGCG 13763
17382 -----GGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17405
13764 CTGCTCGGACAGCG 13823
17406 GTTCTCCCGGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17465
13824 GGTCTTTCG 13883
17466 GAGGTTTCG 17525
13884 CT 13943

17526 GCTGAGGCGCGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17555
13944 GCGCGCATGCTTCCAGCAGACCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14003
17556 GCGCGCGCATGCTTCCAGCAGACCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 17615
14004 CTTCCACCGCTCTCTTCCAGCAGACCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 14063
17616 GCT---GGCGAGGTCTGCGAGCAGCAGCGGTGAGCGCGCGCGCGCGCGCGCGCG 17672
14064 CTTCCGCGAATACCG 14123
17673 GCGCGCGCATGCTTCCAGCAGACCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCG 17732
14124 CTTCCATACCG 14180
17733 GCTGCTACCGTGTGCGAGCAGTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17792
14181 CTTCCACACCG 14240
17793 CTTCCCTCGCGCTTCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17852
14241 CATCGCGCGCATTAACCG 14300
17853 GATCGCGCGCGTAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17912
14301 ACACATACCG 14360
17913 AGAGTCTGCTTCCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17972
14361 CGCTTTCCACTTCCCG 14420
17973 CGCTTTCCAGCG 18032
14421 CTTCACTTACCG 14480
18033 GCTCAGCG 18092
14481 CTTCAACCG 14540
18093 CGAGCGCGTGTCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18152
14541 CCAACCGCTTCCACCG 14600
18153 CCGCGCGGTGAGACCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18212
14601 CACCGCTTACCG 14660
18213 CCAACCGCTTCCAGCG 18269
14661 CCG 14720
18270 CCG 18329
14721 GCACCG 14780
18330 CTTGACCG 18386
14781 CCG 14819
18387 CCG 18425

RESULT 5

US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.

Db	14310	TCCTCTCCGACGGACCCCTCCCGTTCATGACGACGAGAGAGCTGGAAACAGTACTGCGGC	14369
Qy	10940	CCAAGGTGGATGCCGCTCTTTTGTGATGAGCT-----GACCGGGGTATGGAGCTGT	10993
Db	14370	CCAAGGTCGACGCGCGTCTCTCTGACGAACTCACCCTCGACGCCCGCATACGACTGG	14429
Qy	10994	CGGCTTCGTGCTGTCTCTCGGCGCGGGGATCTCTGGGGTGGCGGCGACAGGCAACT	11053
Db	14430	CAGCGTTCGTATGTTCTCTCGCGCGCCCGCTTTCGTGGCGGGGCGAGGGCGCT	14489
Qy	11054	ACGCGCGGCAATCCGCTCTGGACGCGTGGGTACCGCGCGGGCGGGGCTGTC	11113
Db	14490	ACGCGCGGCAACGACCCCTCGACGCCCTTCGCTGGCGCGCGGCGAGCGGACTCC	14549
Qy	11114	CGGGGTGTGCTGGCGTGGGGCTGTGGGAAGAGCGACGCGGATGACCGGCACTGG	11173
Db	14550	CGCCCTCTCTCGCTGGGCTCTGGGCCGAGACGAGGATGACCGGAGTCTG	14609
Qy	11174	CGGCAACGACACCGCGCATCATCGTTCCGGTCTGCAATCCATGTGACCCCCGAGC	11233
Db	14610	GCCAGGCGGACCTGCGCGGATGACCGCGCGGCATCGCGGATCAGCGACCGGAG	14669
Qy	11234	CATGGCCCTCTCGATGGCGCCTGCTCTGACCGGCGGCTCTGCTGCCCGCGACC	11293
Db	14670	GATCGGCTCTCTGACCGCGCCCTCCGCAACGACCGGCAACCGGCTCTGCTGCCCTGC	14729
Qy	11294	TGGCTCCGCGCGCCCTGCG-----	11316
Db	14730	GGCTGAGCGCGCGGCTGGGGAACGCGGCGGGAACGCGCGGATCCCGGCGC	14789
Qy	11317	-----CCCCCTGTGAGGACTCTCTGCGCGCACCGCGCGCGGCAACCCCGC	11363
Db	14790	TCCTTCGGGACGCTCGTGGCGCGGACGCTCGGCGCGCGGCTCGCGGCTCGCCCT	14849
Qy	11364	CACACACACTACCGTGTGGGACACGCG-----	11394
Db	14850	CGACGACGCGGGAACCGCGGACGCGCGGGAACGCGGCGCGGGAACCGCGG	14909
Qy	11395	--GCCAGCTGACGCGCGGCTGGCGGCGAGACACAGAAACAGCACACACCTCC	11452
Db	14910	CGGTACGCTTCGCGACCGGCGCGCACCCTGAGCGGGCGGACGCGGCGCTGCTGC	14969
Qy	11453	TCGCGCTGTGCTCCACATCGCACCGCTCTGGGCGCACACACCCCGGACACCATCC	11512
Db	14970	TCGAGTTCGTGCTCGCGAGGTGCGGAACTCTCGGCGACGCGCGGCTACCGGATCG	15029
Qy	11513	CCCCGACCGCGGTTCCGCGACTCGGCTTCGACTCCCTCACCGCGGCTGCACTAGCA	11572
Db	15030	ACGCGGAACGGGCTTCTCGACCTCGGCTTCGACTCCCTGACCGCGCTGAACTCCGCA	15089
Qy	11573	ACCGGCTCTCCCGACACACGAGACTCGGCTCCCGACACCTCGGCTTCGACCCCA	11632
Db	15090	ACCGGCTCACTCGCGGCTGGCTTGCCCTTCGCGGACCTGCTTCGACCAACCA	15149
Qy	11633	ACCCCAACCTTACCCACCTCCACACAACTCC-----	11671
Db	15150	GCCCGGCGCACTCGCTCCACCTGAGCGCGAGTGGCGCGGCGCTCGGACCAAG	15209
Qy	11672	-----AGCACACCGGACAGCTGTGTC-----	11696
Db	15210	ACGAGCGGGGAACCGGAACGGAACGGAACGGAACGGAACGGAACGGAACGGAAC	15269
Qy	11697	-----CCCGGTGGGGAGCTCGACAACTCGAACTCGGCTCTCGGCTTCGAC	11749
Db	15270	AGACGACGGCTGTGCGACAACTGACCGCTTGAAGGCGCTTGTGCTGACGGGCC	15329
Qy	11750	AAACGACACGCGC---AGCGAAAGTCACTCTCGGCTGAGTCACTCATGTTGAGG-	11805
Db	15330	TCTCGACGCGCCCGGAGGAGAGTCTCGGAGCACCTGCGGCTCCCTGCGCTCGATGG	15389
Qy	11806	-----TGAAACGACCCCGACATCGGACGCGGAGGAGTGGGGATGCTCTGCTGT	11845
Db	15390	TACGCGGCGAGACCGGGAACCGGGAACCGGTCGGAAGCCCGCGGACGCGGCTCGGCG	15449
Qy	11846	ACGACGA-----	11852
Db	15450	CCGAGGACCGGCCCTGGGCGCGCGGACGAGCGGGGCGGAGTGAGGACGCGCGG	15509
Qy	11853	-----GAAGTTCATCATCGGCAACAGAGCTGAGATTTTCAAATTCATTGACAGAC	11905
Db	15510	GAGTGCAGGACTTCATGAACGCTCGGCGGAGAACTCTTCGCGCTCTCTCGACGAG	15569
Qy	11906	TGGGCTGTCTTGAACCGGACGCTTCGCACTCGGCGGCTATCG-----	11949
Db	15570	CCAGCAGGACTGATCTTCGCGCAGCGTTCGCTCCGCGCGGACCCCGTCCCGGCGAC	15629
Qy	11950	-----CTGGGCGCTCTGTAGGACGTGAAT	11972
Db	15630	CTCGACTCGAATCACTTCATGCGGCTCTCGGCGCTCCAGGAACTCAAGGGGACAGCGT	15689
Qy	11973	GCAATTTGGCAATGAAAGCAAGCTCTCTGGAATACCTCAAGCGGCTCATGCGGAC	12032
Db	15690	GTCCAGGTGAACGAAAGATGACTCTGACTACCTGCTGCTGCGCACGCGGACCTCCA	15749
Qy	12033	CGGCACTCGCGCTCGCTGTAAGAGTGTGAGGCTGAGCGTGAGGAGGCGGATCTG	12092
Db	15750	CGAGGCGCGTGGCGGCTCCGCGAGTGGAGGAGGCGGCGGAGCGGCGGATCGT	15809
Qy	12093	GGGGATGGCGTGTCTGTTACCCAGGCGGGCGACGTCACCCAGCGGACTGTGGCATCTCGT	12152
Db	15810	CGGATGGCTGCGCGCTGCGCGGCGGCTCGCTCGCGCGGAGGACTGTGGCGGCTGT	15869
Qy	12153	CAAGTCCGACGACGAGCTATCGGGAGTTCCCGACCGACCGTGTGATGAACTGAGACA	12212
Db	15870	GGCGCGCGGAGGACGCGGATCTCGGAGTTCCCGGAGGCTCGGAGCGGCTGGGAG	15929
Qy	12213	GCTCTAGACCGGACCGCGGCTCAGGAACGAGTTACACCGCGGAGGAGGTTCT	12272
Db	15930	CCTGTACGACCGGAAACCGGAGGCGACGCGGAGGATTCGCGCGGAGCGCGATCTCT	15989
Qy	12273	CTATGACGCGGCGACTTCGACCGCGGTTCTTCGAGTTGTCCACCGCTGAGGCGCTGCG	12332
Db	15990	GTAAGGCGGCGGATTCGACCGCGACTTCCTTCGCGGATCTCGCGCGGAGGCGCTCGC	16049
Qy	12333	AATGACCGCGACGAGCGCTCTCTCGAAACCACTTTGGGAAACGTTGCAACAGGCGG	12392
Db	16050	CATGACCGCGACGAGCTCTCTCTGAGGCTCTCTGGAGGCTCTCGGAGCGGCTCGAG	16109
Qy	12393	AATCGACCGGAGTCCATGCGGAGGCGGACCGGCGGTTTCGTTGGGATCAATCGGA	12452
Db	16110	GATCCCGCGGCGCACCGCGCGGCGACCTCGGTCGCGGCTCTTACCGCGGCTGATGACCA	16169
Qy	12453	GGACTACACACCGGATACACATCAGCCCTCAAAACGAGTCGAGGCGTACTGCTCAC	12512
Db	16170	CGACTACGCGACCTCTCTCACCGAT---GTCCCGGAGGCGATCGAGGCGTACTGCGG	16226
Qy	12513	TGGAGCGGCGCAAGATGCGCTGAGCGGCTATCTCTACAACTTCGGGCTCGAAGGCGC	12572
Db	16227	CGGCAACTCGGCGAGTGTGCTCGGCGCGCTCGCTGAGGCTGAGGCTGCGGAGGCGC	16286
Qy	12573	TGCGATCACTATCGACGCGGTTCTCTCTCGCTCTCGGCTCTGCGCTGCACTGCGCTG	12632
Db	16287	GGCCGCTACGCTGCGACCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	16346
Qy	12633	AGCGCTCGGCTCGGATGAAATGACCATGCGCTCGCAGGCGGCGCTCTCTCTCTCTCTCT	12692
Db	16347	GGCCCTGCGAAGGCGGAGTGCATGCGGCTCGCGCGGCGGCTGACGCTCATGCTGAC	16406
Qy	12693	TCCCTTCGTTTACCGGAGTCTCTCTCGCAGCGGCGCTGCGCGGAGCGGCGGCTGCA	12752
Db	16407	GCCAGCACCTTCGTCGAGTTGAGCGGCTGAGCGGCTGCGCGCGGCGGCGGCGGCTG	16466
Qy	12753	GCGGTTTTGGGCGGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	12812
Db	16467	GTCTTCTCTGCGACGCGCGGCGGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT	16526

Db 10025 ACCTTCGGTCTCGAAGGCCCGCGACGACCGTTCGACACCGCCTGCTCGTCTCGCTGAC 10084
Qy 6790 GCTTTTCATCTGCGCTCTCAGGCGTTGCGGTGCGGTGAGTCTCGATGCGCTTCGCGG 6849
Db 10085 GCCTTGACCTGGGGTGGGGCTGCGAGCGGAGTGACGATGGGCTCGCGGT 10144
Qy 6850 GGTGTACGGTGTATGCTCTCGGGTGCCTTCGTGGAGTTTTCGGGAGCGGGGTCTG 6909
Db 10145 GCGGTGCGATGATGCGACCCCGCACATGTTCTGTGAGTTTCAGCCGTACGCGCGCTC 10204
Qy 6910 GCCCGGACGGGCATTTGCAAGGGTTCGCGCGCGCGGACGGGACCGCTGGGTGAG 6969
Db 10205 GCCCGGACGGCGCGACGAGGCTTCTCGCGGACCGCGAGGTTTCGGCGCGCGGAG 10264
Qy 6970 GGTGTGGGATGCTGCTGTGTGAGCGGCTCTCCGAGCGCCCATCGCAACGGTCAACGCTGTC 7029
Db 10265 GGGTCTGGCTGTCTGTGTGAGCGGCTCTCGAGCGGGGCGCAACGGTCAACCGGTG 10324
Qy 7030 CTGGCGGTGGTGGGCAAGTGGGTCAACGAGGAGCGTTCGAGCAACGGTCTGACCGG 7089
Db 10325 CTGCGGTGGTCCGGGTACCGCGGTCAACGAGGAGCGGCGAGCAACGGGTGACCGG 10384
Qy 7090 CCCAAGGGCGGCTCCAGAGCGGTTCATCCGCGAGGCGCTCGCCAAACCGCGCTGTG 7149
Db 10385 CCCAAGGGACCTTCGAGGAGCGGGTTCGCGGAGGCGCTCGCCAAACCGCGCTGGCA 10444
Qy 7150 GCCGGTGTGTACACGGGTGAGGCCACCGGACCGGACCACTTTGGCGGACCGATC 7209
Db 10445 CCCGGGACATCGACCGCTCGAGACGCGACCGGACCGGAACTTCGTGGCGGACCCATC 10504
Qy 7210 GAGGCCAGGCGCTCTCGGACCTACGGACAGGACCGTTCGCGGAGGCGCGCTGTGG 7269
Db 10505 GAGGCCAGGCGCTCCAGGCGACGTACGGCAAGGAGCGCGCGGAAACCGCGCTGCGC 10564
Qy 7270 CTGGGCTCGGTCAAGTCCAAATGTCGTGTACACAGGCTTCGCGGGCGTTCGCGGGGTG 7329
Db 10565 ATCGGCTCGGTGAAGTCCAAATGTCGTGTACACAGGCTTCGCGGGCGTTCGCGGGCATC 10624
Qy 7330 ATCAAGATGTATGCGCTGGGCGATGTCGTCTCGCGGAGCGTTGATGTGATGAG 7389
Db 10625 ATCAAGATGTCCTCGGATGCGCCACGGCACCTTCGCCAAGACCTTCACGCGGACGAG 10684
Qy 7390 CCGTCCCGCATGTGACTGGTTCGCGGGTGGGTCAGTGTGTCAGCGAGACGGTGCC 7449
Db 10685 CCGAGCCCGCACGTGACTGGGCGAACAGCGGCTGCGCTCTGTACCGGACCGATCGAC 10744
Qy 7450 TGGCCCGGGGAGGGGGGCTACCGCGGCGAGGAGTGTATCATTTGGGCTGACGGCG 7509
Db 10745 TGGCCCGGGGACCGGTCG---CGCCGCGCGCGCTCTCTCTCTTCGGCATCAGCGG 10801
Qy 7510 ACCAAGCGCCAGTCTCTGAAGAGGACCGCGCGAGCGT--- 7554
Db 10802 ACAGCGCGACGTGCTGTGAGAGGCGCGCATGCTGTGTGAGTGTCTGGGGCC 10861
Qy 7555 -----CCGCGGGGACCAACC 7569
Db 10862 GATAGGTGCTGAGTGTCTGAGAGGTAGCGATGGCTGGGACGGCTGGGACCTCGAG 10921
Qy 7570 GCGCGGAGGGTGAACCGCGGACGACGATGAGGTGCTGCGGAGTCTCTGGGGT--- 7627
Db 10922 GTGCTGAGGGTCTGAGGCTCTCGAGGCTCCGAGGCGCGCGCGCGGACCGCTGAGGCGTCC 10981
Qy 7628 -----GGCGGTGGTGTGCGGCGGAGTTCGAGCGGCGCTGCGCGGCGGAG 7674
Db 10982 CTCGCCGGGACCTGCGCTGGGTGTGTCGCGCAAGGACGAGAGTGTGCGGCGGAG 11041
Qy 7675 GCCCAGGCGCTGACGCGCCACCTCACCGGACCGCGGCGCTCGA----- 7718
Db 11042 GCGCGCGCTGACGCGGTGGTGTCCGAGCGCGCGCGCGCGGACCTGTGCGGACGGA 11101
Qy 7719 -----CCTCGGAGTGTGGATACACCTTCGCGGCGCGCGCGGCGGTGTGACACAC 7770
Db 11102 CCGGCGCGCGCTGCGGAGCTGCGGTACAGCTCGCCACGAGCGGTACCGCTTCGCGCAC 11161

Qy 7771 CCGCCACCTCATCGCGGAGACCGCGACAGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 11162 CCGCGCGCGTGAACCGCGCGACCGGACGGGTTCTTGACGGGTGCGCACGCTGCGC 11221
Qy 7831 GCAGGAGAGCCCAACCGCGCGCTCATCACAGAGGCGCCCGGGCGGAGACCGGAGCCGGG 7890
Db 11222 CAGGGCGGACCTCGCGCGACCTGACCTGGAACACCGCGCGGACCGGAC----- 11272
Qy 7891 GAGGCGGAGGAAAGACGCACTTCCTCCGAGAGGCGACCAACGCGCGCGCATG 7950
Db 11273 -----ACCGGTTCTTTCACCGCGAGGCGAGTACGCGCGCGCGCGC 11317
Qy 7951 GCGCCAGGCGCTTACACACACCGCGCTTCCTGCGCGCGCACTCAACAGCATCTGCACC 8010
Db 11318 GCGCGTGTGCTGACGACCGGACCGCGTTCCTGCGCGCGCTCGACGAGATCTGCGC 11377
Qy 8011 CACCTCGAGCCCACTCGACACCGCTCTCTCCCGCTCTCACCCAAACAGCAACGAC 8070
Db 11378 CACCTCGAGCGTCACTCGAACTGCGCTCTCGACGTAT-----GTTCCGCGCGGAG 11431
Qy 8071 AACGAGGAGCGCGCGCACTGCTCCAGAGACCGCTAGCGCGCGCGCGCTCTTCGCGC 8130
Db 11432 GGCAGCGGAGGCGCGCTGCTCGACGAGAGCGGTCACGAGTGGCGGTGTTTCGCGC 11491
Qy 8131 TTTCAGGTGCGCTCCACCGCTCTCACCGAGCGCTACCACTCACCGCGCGCTACTAC 8190
Db 11492 CTGAGGTGCGCTCTTCGCGCTGTC---GAGAGTGGGATGCGGCGCGCGCACTG 11548
Qy 8191 GCGGACACTCTCTCGGCGAAATACCGCGCGCACTCGCGCGCATCTCTACCTCTACC 8250
Db 11549 CTCGGTCACTCGGTGCGGAGATCGCGCGCGCACTCGCGGTGTGTTCTCGCTGCGC 11608
Qy 8251 GAGCGACCACTCTCATACCAACCGCGCGCACTCATGCAACCATGCGCGCGC---GCG 8307
Db 11609 GAGCGCGCGCTGCTGCGCGCGCGCGCGGTCTATGCAAGAGTGCCTCGCGCGGTGC 11668
Qy 8308 ACCATGACCACTCTCCACACCGCGCGCGCGCGCGCTACCGCACTCACCGCGCGCAAA 8367
Db 11669 GGTGCTGCGGTTCAGGCGCGGAGGACGAGATCGCGGTGTGCTGGAGACGAGGAG 11728
Qy 8368 AACGAC-----CTCGCATCGCGCGCACTCAACACCGCGCGCTCTCTCTCATACG 8418
Db 11729 CGGTACGCGGAGCGTCTGACGCTCGCGCGCTCAACGCGCGCGAGCGCGCGCTCTGTC 11788
Qy 8419 GGCACCGCGCACCGTCCACACATCACACCTCTGCGCAACACAGGAGCATCAAAAC 8478
Db 11789 GCGGAGCGGAGCGCGCGCGGAGCGGCGGTACTGCTCGGGCTCGGCGCGGAGAC 11848
Qy 8479 AAACCTCTCCCAACACCGCGCTTCACCTCCCGCGCGCGCGCGCGCTCTCAACCAA 8538
Db 11849 GCGCGCTGCGGTCAGCGCTTCACCTTCGCGGACATGAGAGGATGCTCGACGGG 11908
Qy 8539 CTCACAGGACACCAAACTCTACCTACCAACCGCGCGCGCGCGCTCTATCAACCGCGC 8598
Db 11909 TTCCGCGCGCTCTGAGAGCGGTGAGTTCGCGCGCGCTCTCTGACCGTCTCGAAC 11968
Qy 8599 AACA-----CCCCACCGAACCACTCTACCGCGCGCGCTCTGAGACCGCAAC 8646
Db 11969 GTCACCGGCTGCGCGCGCGCGCGCGCGCTGTCGAGCGCGCGGAGTCTGGGTTCGCGC 12028
Qy 8647 GCGCGAACCGCTGAGTACGCGACCGCGCGCGCGCGCGCGCTCCACCAACCGCGCTCAC 8706
Db 12029 GTCCGCGGACCGCTCGCTTCTCGACGGGCTCGGTCTCTGCGGAGCTCGCGGTGCGG 12088
Qy 8707 ACCTATCGAAGTGGAGCCGAGCAACCTCTACACCTCTACCGCGCGCGCGCTCTCC 8766
Db 12089 ACCTGCTGGAGTGGGCGCGCGCGCGCGCGCGCTCTACCGCGCGCGCGCGCGCGCTGCGG 12148
Qy 8767 AACCG 8799
Db 12149 GACACCGCGCGGATTCGCGTTCGCGCGCGCGCGCGCGCGCTCTCTCGCGCGCGCTCTCCCGCGC 12208

Db 6047 GGCACGGGCGTGACGACGAAACAGCCGCTCGGCTCGGTCGTTGAAAGTCAAACATCGGG 6106
Qy CACACCATGGCTGCGCGGGGTGTGGGTGGGTGATCAAGATGGTGTGATGGCGTGTGCGGGAG 2550
Db CACACCATGGCGCGGGCGGGGTCTCGGTGTATCAAGATGGTGTGATGGCGGTGTGCGGGAG 6166
Qy GGGGTGTGCGGGGACGTTTGTATGTGATGTAAGACCGTTCGCGGAGGTGATGATCGGTTCGGG 2610
Db GGAAGTGTGCGGAGACGCTGCACTGACGAGCCCTCGACACAGATCGACTGGTTCGGCT 6226
Qy GGGGCGGTGCGGCTGTGACGAGGCGGTGCGGTGCGGCGGGGACGCGGAGGGCGGTG 2670
Db GGGGCGGTGAAACTCTCACGAGGCGTCACTGCGGCGGAGAGAGAGACGCGGCGGTG 6286
Qy CGCGGCGGAGTGTCTGCTTCGCGATCGCGGCGACGAATCGCATGTGATTTGGAG 2730
Db CGCGGCGGCGGTCTCTCTTCGCGATCAGCGGCGACCAATCGCATGTGATTTGGAG 6346
Qy GAGGCGCGGCGGCGGCGGTGTGTTGCGGCGGCGGTGGGTGTGAGAGGTGCTTCGCGGT 2790
Db GAGGCGCGGCGGTGTGCGAGGTGTCTCGGTGCTCGAGCGGTGCGGTGCGGCGGTGCGG 6406
Qy CTTGCAATTTGCGGTGAGTGTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2850
Db GTGCGGCGGCGGTGTGACGCGCTTGGGTGTGCGGCGGAGGTCCGCTGCGGCGGTTCGAGCGG 6466
Qy GAGTGTGCGGCGGTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2910
Db CAGATCGAGCGGT-----TGCGCATTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6503
Qy GCTGGGTGCGGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2970
Db GTACGGATGACGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6537
Qy TCGCTTGCAGATGTGGGTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3030
Db GGGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6538
Qy GTGCTGCTGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3090
Db GTGCGGTGCGGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6593
Qy GAGCCGATGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3150
Db CTGATACGCGGAAACGCTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 6688
Qy TTGCTGTTTCCGCGAGAGGTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3210
Db TTGCTGTTTCCGCGAGAGGTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6748
Qy TCTCCGCTGTTCCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3270
Db TCGCGGCTGTTCCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 6808
Qy TGTGCTGTGTGAGCATCTGCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3330
Db TGTGCTGTGAGCATCTGCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 6868
Qy GTGGTCCAGCTGTGCTGTGAGCATCTGCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3390
Db GTGCTGAGCATCTGCGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 6928
Qy GGTATCGAAACCGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3450
Db GGTGAGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 6988
Qy TGTGGGCGGTGAGCTGAGGAGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3510
Db GCGGAGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 7048
Qy GCGGCGGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3567
Db GCGGCGGCGGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 7108

Qy 3568 GAGCAGCTCATTTGTTGAGCGGTGCGGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3627
Db CTTGAGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 7109
Qy CCGTCCACCGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGT 3628
Db ACCGCGCTGCTGCTGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGT 7163
Qy GGCACCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCG 3688
Db GCGGACCGGATTCGCGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCG 7223
Qy CAGCCCTTCGCGGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 3748
Db GAGATCATCGAGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGTGAGGAGTGT 7283
Qy GTGCGGTTCCTTCCACCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTG 3808
Db GTGCGGTTCCTTCCACCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTG 7343
Qy TACTGTTACCGAACTTCGCTCAGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCG 3868
Db TACTGTTACCGAACTTCGCTCAGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCG 7403
Qy GATGAGCGGACACCGGCTTCTGTCGAAGTACGCGGCGGTGCGGCGGCGGTGCGGCGGCGGTGCGG 3928
Db CGTACGAGGCGTTCAGCGCTTCTGTCGAGGCGGTTCAGCGCTTCTGTCGAGGCGGTTCAGCGCT 7462
Qy GAAGACACCGCGGAGACACCGCGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGTG 3988
Db CCGGCTCTCACCATGACCGCTCCCGGAGCGGTGAGGCGGCGGTGAGGCGGCGGTGAGGCGGCGGT 7505
Qy GACACGACACCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGAC 4048
Db CAGGAGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGT 7565
Qy ACACCGACCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCG 4108
Db GTGCGATGAGCTTCCTGTCGCGGCGGTTCCTGTCGCGGCGGTTCCTGTCGCGGCGGTTCCTGTC 7625
Qy CACCTCGACCTGCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTG 4168
Db CGGTTCGCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCG 7666
Qy CCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 4228
Db GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 7721
Qy GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 4288
Db GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 7737
Qy GACCTGGAACCGGTTCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTC 4348
Db GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 7796
Qy GTGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGT 4408
Db GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 7856
Qy ACCTACCGAGGAACTTCGAAACCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGT 4468
Db GCGGCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCG 7916
Qy TGCGTTCATCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGAC 4528
Db TGCGTTCATCGGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGAC 7964
Qy AACCTCCACCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCG 4588
Db AACCTCCACCGCGGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCGCGGTTCCTGACCG 7964

8465 GCGAACAAGCCCGCGAGCCACCACTCACCGCGAACTACACCGATCGGCGCCCGC 8524
5308 CTACCATACCACTCGGACACAGCAACACCCAGACCAACTCCAACTCTCTCAACACC 5367
8525 GTACCATCGCGCTCGAGGTGCGGACCCCAAGCATGCGACCTCTCTGAGCC 8584
5368 ATCCCGCCCAACACCCCTCTCAACCGTATCCACACGCGAGCATCTCTGACGAGCC 5427
8585 ATCCCGCGGAGACGCCCTCACCGCGTGTCCACACGCGCGCGCTCGAGACGCC 8644
5428 ACCTCACCACTCAACCCCAACCACTCAACAAAGTCTCTCGCGCAAGCCACAGC 5487
8645 ATCGTGACAGCTGACCGCGAGAGGTCCGCGCGCCACCGCTCGGAAGCGCTCGC 8704
5488 GCCACCTCTCTCACCACTCACCAACACACACCCCTCTACCGCTCTGCTCTACTCC 5547
8705 GCCTCGGTCTGACAGCTGACCGGGAACCTCGACTCGACGCTGCTGCTCTCTCG 8764
5548 TCCGCGCGCCACCTTTCGCGGACCCCGGCAAGCCAACTACGCGCAGCAACGCTTAC 5607
8765 TCCGCTGCGAGCACTCTGGGCACTCCCGGTTCAGGGCAACTAGCCCGGCAACGCTTAC 8824
5608 CTGACGCGCTCGCCACACCGGCAACCCACCACTCCCGGCAACGAGATCGGCTGG 5667
8825 CTCGACGCGCTCGCGCTCGCGCGCGGCGCACCGGCGGCTCGCGCTCTCGGTGGCTG 8884
5668 GGCACCTGGCAAGAAACGACTCGCTGTATTCGGACAAAGCCGCGCATATCTCGACCGC 5727
8885 GGCAGTGGACGCTGGCGGATGCGCGCGGCTGCGCGGTGCGGCTGCGGCTGCGCAAC 8944
5728 CGCGGTTTGCACCCATGTACCCGAGTTGGCCACCGGACG----- 5768
8945 CACGGCTGCGCGCATGACCCGGAACCTCGCCCTGGCGCACTGGAGTCCGCGCTCGC 9004
5769 -----GGTACAGGGGATCGCGGACACCGAAGCGGCTATGCTATCGCGGAC 5820
9005 CGGAGCAGACCGCGATCACCGTTCGGGACATCGACTGGGACCGCTTCTACCTCGCGTAC 9064
5821 ATCGACTGGAGCAAGATCGAACAACCTCTGACACGAGCACTG-----GTGAGCGG 5874
9065 TCTCGGTGCGCGCGAGCTCTGTCGAGGAGCTCCCGAGTGGCGGCACTATCTGAC 9124
5875 GCCCGGGAAGGAGCAGTGTCTCAGCGCCCACTTCCACCGGCGG-----AGTTG 5925
9125 GCAAGGACAGCGCCAGCTCCGAGACGCGGAGCTCCGCCAGGCGGCAACCCCTG 9184
5926 CACAAACGCTGCGCCATCAGAGCTCGCGGACCAAGCGGCGCATGTCTGAGCTGTA 5985
9185 GCGAGCGCTGGCGCGCGGCTCCCGGAGCGGTACGGAGATCTCTCTCGGTCTCGTA 9244
5986 CGAGACATGTGGCGGAGTGTCTCGGCAACCGGACCCGAAAGCCATCGCGCGGACAG 6045
9245 CGGCGCAGCGCGCGCGGTCTCGGATGCTTCGGAGAGTTCGCGCGGAGCGTCCGCGGACCG 9304
6046 TCGTTCGCTGCACTCGGCTTCGATTCACCTCAGCGCGCTGAGTTCGGAACCTGTGATC 6105
9305 GCCTTCAAGGACATCGCTTCGACTCGCTCGCGGTGTGAGTTCGCGCAACAGGCTGACC 9364
6106 AAGGCAACAGGACTCGCGCTTCTGTCTCGTGTCTTTCGACCAACCGGACCCCTGCGAAA 6165
9365 CGGCGACCGGCTCAGGTGCGCGGAGCGCTCTTCGACCAACCGGACCGCGCTGGCC 9424
6166 CTCGCGGTACCTGCGAGAACCACTGCGGGGACAGAGCGAGTCCG----- 6215
9425 CTCGTGCTGCTCGCAGAGTTCCTCGGTGACGAGAGAGCGCGGACCGCGCGGCGG 9484
6216 -----TCCTTCAGCGGCGAGCGGTTTACCGCGCGAGGCTTCT 6249
9485 TCCGCGCGCTGCGCGGACTGTGCGGTGCGGCTGCGCGCGCGCGCGCGGCGGATGCC 9544
6250 GTACAGCGCGATCGCCATCGTTGGCATGGCTGTGCTTTCGCGCGGAGTGACTCG 6309
9545 GACGACGATCCGATCGCGATCGTGGGATGAGTGGCTGCGCTACCCCGGTGACATCGCAGC 9604

6310 GCGGACGACTTCTGGGATCTGATCTCTCCGAGCAGGACGCGATCGCGGATTTCCCCACC 6369
9605 CCGGAGGACTGTGCGGATGCTGTCGAGGGCGGAGGAGCATCACGCGGTTCCCCACC 9664
6370 GACCGCGCTGGGACTGTGACAGCTCTAGACCCCGGACCCCGACACCCCGGCACTGCG 6429
9665 GACCGCGCTGGGACTGTGACAGCTCTGACGCGCTGTAGACCCCGGACCCCGGAGGCG 9724
6430 TACACCCGAAACGCGGATTTCTCTACGACGAGGACCTTCGACGCGCAATTTCTCGGC 6489
9725 TAGTTCGCGAGGGGCGGTTCCTGACGACGCGGCGAGTTCGACGCGAGTTCCTCGGC 9784
6490 ATCAGCCCCCGGAAAGCCCTCTCGCATGGAACCCCGAGCAACGACTCTCTCGAAACCGCC 6549
9785 GTCTCGCGCGGAGCGCTGCGCATGGACCCCGAGAGGAGTGTCTCTGACGAGCTCC 9844
6550 TGGGAAACCATGAACACGCGCGCATCAACCCCGACACCTTCACGCGCACCCCGACCGGA 6609
9845 TGGGAGGCTTCGAGCGGCGCGCATCGAGCCGCGATCGTTCGCGGAGCAGCAGCAGT 9904
6610 GTCTTACCGGACCAACGAGCAGACTACGACTTCGCGTGCACACCGGCGGCGGCTCA 6669
9905 GTCTTATCGGCTCTCTTACGAGACTACGCGGCGCGCTCCGAAACCGCGCGTGGC 9964
6670 ACCGATGCTTCGCACTGACCGGAAACCGCGGCGGCTCATCTCCGCTGCTATCTCTGATC 6729
9965 GTGAGGCTTACTGCTGACCGGAGCAGCGCGGAGCTCGCGTGGCGGCTATCGGCTAC 10024
6730 AGTTTGGTTTGAAGGCTCTCGGTGTCGGTGAACAGCGTGTTCCTCGTGTGTTG 6789
10025 ACCTTCTGCTCTGAAAGGCGCGGACCGCTCGACACCGCTGCTGCTGCTGCTGAC 10084
6790 GCTTTGCTATCGGCTGTGAGGCTGCGTGGGCTGAGTGTGCTGAGTGTTCGCGGCGGCTG 6849
10085 GCGCTGCACTGCGCGTGGCGGCTGGCGAGCGGCGAGTGCACGATGCGCTGCGCGGT 10144
6850 GGTGTGACGCTGATGCTGCTCGGCTGCTTCTGTTGAGTGTTCGCGGCGGCGGCTCTG 6909
10145 GCGTGGGATGATGAGGACCCCGCACATGTTCTGTTGAGTTCAGCGCTCAGCGGCGCTC 10204
6910 GCGCGGAGCGGCAATTCGAAAGCGTTCGCGCGGCGGAGCGGACCGGCTGGGCTGAG 6969
10205 GCGCGGAGCGGCGGAGCGCTTCTCGCGGAGCGCGGCTTCGCGCGCGGCGAG 10264
6970 GGTGTGGGATGCTGCTGTTGAGCGGCTCTCGAGCGGCTTCGCAACGCTCACCTGTC 7029
10265 GCGCTGCGGCTGCTGCTGAGCGGCTCTCGGACGCGCGCAACGCTCACCGGCTG 10324
7030 CTGGCGCTGCTGCTGGGAGTCCGCTCAACAGGACGCTGCGAGCAACGCTCTGACCGCG 7089
10325 CTCGCGTGTGCTCGGCTACCGCTCAACAGGACGCGGCGGAGCAACGCGCTGACCGCG 10384
7090 CCAACGCGGCGCTCCAGCAGCGTGTCTATCGCGGCGGCTTCGCAACCGGCTTGTGCG 7149
10385 CCAACGAGACCTTCGAGCGGCTGATTCGCGGAGGCGCTCGCGAGCGCGGCTGGCA 10444
7150 GCGGCTGATGTCAGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGATC 7209
10445 CCGGCGGAGATCGAGCGGCTCGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10504
7210 GAGGCGGCGGCGGCTCTCGGACCTTACGAGACGAGCGGCTGCGCGGCGGCGGCGGCTG 7269
10505 GAGGCGGCGGCGGCTTCAGGCGGCTGACGCGGAGGAGGCGGCGGCGGCGGCGGCGGCG 10564
7270 CTGGCTCGGTCAGTCCAAATGCTGCTCACACAGGCTGCGCGGCGGCTGCGCGGCGG 7329
10565 ATCGGCTCGTGAAGTCCAAATCGGACACACCGGCGGCGGCGGCGGCGGCGGCGGCGG 10624
7330 ATCAGATGCTGAGCGGCTGCGGCGGCTGCTGCTGCGGCGGAGCGGCTGATGATGAG 7389
10625 ATCAAGATGCTGCTCGGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10684

Db 4506 TGGCTGCGGACAAAGCGGGTACGACCGCTAGTCGCTCCCTCGCGGAAGCATGGGCCA 4565
Qy 971 CCGCCGACACCGGTGATCATGGGACAGCTGCGCGCGGCCAGAGGCACCTTGAGACCATCTCC 1030
Db 4566 ACGGACTCGGGTGAAGTGGAGCCGCTCTCCCTCGCGACCGGCGACCATCTCGACC 4625
Qy 1031 TGAAGTCTCTCGC-----CQAATACGGGGGACATGTGAGACGTGCGGCA 1075
Db 4626 TCCCCACCTACCGGTTCAGACCGAGCGCACTGGGTGGCGAGATCGAGGCGCTCGCCC 4685
Qy 1076 CACCGTCTCTCGGACAGCGCTGACCGCGCTGTCGCCACGAG----- 1119
Db 4686 CGCGGGCGAGCGCGGGTGCACCCCGCTCTCCGCAAGGCGGCGCGAGCCCGCG 4745
Qy 1120 -----CAGCAGTCTGCTCTGACCTGGTGGCGGCCACACCATATGGCGGTCTGAACG 1174
Db 4746 AGTCTGACCGGACGAGCAGCTGGGTGATCTTGACAAAGTTCGGGCGCAGACGGCCC 4805
Qy 1175 ACGACGGAAACGAGCGCACCGGTGCGATGCCGCCCATCGGCGAGTTTCGCCACCTCG 1234
Db 4806 AGGTGCTGGGGTACGCGACAGCGGGCGAGATCGAGTTCGACCGGACCTTCGTTGAGGCG 4865
Qy 1235 GCTTCACTCCGTATGGGTGTCGAATCGGCAACCGCTCAGCAAGGCGACCGGCGCTCG 1294
Db 4866 GTTGCACTCCCTGACCGCGGTGGACCTTCGCAACCGGATCAAACGCGCTTCGCGGTAC 4925
Qy 1295 GGTTCGCGGTGACGCTCATCTTCGACACACACCGCGCGGGTTCGCGCGCGCTTC 1354
Db 4926 GATGGCGGCTCATGATCTTCGACTTCCCAACCCCGAGGCTCTCGCGGAGCAGCTCG 4985
Qy 1355 GGAACCGCGGCTTCGCCACCTCGACGAGGACACCGCGCCCGTACCGGCTACCCAGCG 1414
Db 4986 TCCTCTGCTGACGCG-----GGAGGCGGGCGGAAACCGCGCGGTGCGGAGCGGCT 5038
Qy 1415 GCGACGAGGC-ACGCGAGGGCGGACGACCGCGATCGCATCATCGGATGCGATCGCT 1473
Db 5039 CGGTGGCGGGCGCGGTGCGTGCAGCGAGCGCGGTGCGATCGTCGCGATGCGCTGCCG 5098
Qy 1474 TTCGGGGCGGAGTCCGCTCCGGAAGGACCTGTGGAGCTGCGCGCTCGGCGGAGAC 1533
Db 5099 CTGCGCGTGGGTCCCTCGCGGAGGACCTGTGGCGGTGTGGCGCGGGCGGGGAC 5158
Qy 1534 GCGATCTCGAGGTTCCGCGAGGACCGCGCTGGGAGTGGAGGGGTGTACCAACCGGAT 5193
Db 5159 GCGATCTCGAGGTTCCGCGAGGACCGCGCTGGGAGTGGAGGGGTGTACCAACCGGAT 5218
Qy 1594 CCCAGCGCGCGGACGTTCTATCCGAGGAGGCGGGTTCCTTCAGACGCGGGCGAC 1653
Db 5219 CCGGAGCACCCCGGCACTGTCGTCAGCGCGCGGTGCGATCGTCGCGATGCGCTGCCG 5278
Qy 1654 TTCGACGCGGCTTCTTCGGAATCAGTCCAGCTGAGGCACTGGCGATGGATCGGAGCAG 1713
Db 5279 TTCGACGCGGCTTCTTCGGAATCAGTCCGCGCGAGGCGCTTCGCAATGACCCGAGCAG 5338
Qy 1714 CGGCTGCTGTGAGACGCTCTGGAGGCTTCGAGCGGGCGGGAATCGATCCGCTGTGCG 1773
Db 5339 CGGCTCTCTCGAATCTCTGGAGGCGGTGAGGACCGCGGATCGACCGACCTCC 5398
Qy 1774 GTACGCGGTTCGATACGGGCTTTCGCGGGCGCTCTCTTCGATACGCGCGCGGT 1833
Db 5399 CTGCGGGGACGCGGAGTGGCGCTTTCATGCGGGCGATGACCAAGATACGGGCGGAGC 5458
Qy 1834 ATGGACACCGCTTCGAGGCGGCGCGGACGTTGGAGGGCCATCTCTACCGGTACC 1893
Db 5459 CT-----GCGGACGCGCGGGGAGGCGCTCGACGGGTACCTGCTGACCGGCAAC 5506
Qy 1894 ACGGGCAGCGCTCTGCGGGCGGTATCGCTACAGCTTCGGGCTGGAAGGCGCGGATC 1953
Db 5507 ACGGCCAGCGTATGTCGCGCGCGCTCTGTACACACTCGGCTTCGAGGCGCCCGCCCTG 5566
Qy 1954 ACGGTGGAACCGGGTGTGCGGATCGCTCGTGAACGCTGATCTGGCGGTGCGAGTGTG 2013
Db 5567 ACGGTGGAACCGGCTGCTGCTGCTGCTGCTGCGCTGCGCTGCGCTGCGCGCGCTG 5626

Qy 2014 CGGTGCGGTGAGTGCACGCTCGCTGCGCGCGCGGTCTCGGTTCATGTCCACCTTCGGC 2073
Db 5627 CCGAGGGCGAGGTGCATGCGCTCGCGCGCGGTGGCGTGTGATGCCACCGCCGG 5686
Qy 2074 ATGTTTCATCGAGTTCTCCCGCAGCGCGGCTGTGCTGTTGAGACGAGGTGCAAGGCTAC 2133
Db 5687 ATGTTTCATCGAGTTCTCCCGCAGCGCGGCTGTGCTGTTGAGACGCGGTGCAAGGCTTC 5746
Qy 2134 TCGGCTGACGCGACGCGCACCGCTGGGCGAGGGGCTCGGATGCTGTTGTTGAGCGG 2193
Db 5747 GCGCGTTCGCGGAGACGCGACACGAGTGGTCCGAGGGGCTCGGCGTCTCTCTGTCGAGCGC 5806
Qy 2194 TTGTTGAGTGGCTGCGGTGCGGTGCTGGCGGTGATCGCGCGGTGATCGCGCGAGTGGCT 2253
Db 5807 CTGTGCGACCGCCCGCGCAACGGAACACAGGTCTCTCGCGTGTCCGCGGAGCGCTTG 5866
Qy 2254 AACAGGACGCTGCTGCAATGGGCTGACCGCGCGCAACGCTCCGCTCAGAGCGGGTG 2313
Db 5867 AACAGGACGCTGCTGCAATGGGCTGACCGCGCGCAACGCTCCGCTCAGAGCGGGTTC 5926
Qy 2314 ATCCGCGAGGCTGCGGACGCGGGGTGTCTGCGCGGATGCTGATGCTGTTGAGGGG 2373
Db 5927 ATCCGCGCGCTGCGGAGCGCCGCGTACGACCTCCGAGTGGACGCTCTGTCGAGGCA 5986
Qy 2374 CACGCGACGCGGACGACGCTGGTGTATCCGATCGAGGACAGCGCTTGTCTCGCCAGCTAC 2433
Db 5987 CACGCGACGCGGACGCGGACTCGCGACCGCTCAGAGCGGCGAGCGCTGTATCGCCACTAC 6046
Qy 2434 GGGCAG---CGGGCGGTGACAGCGCGTGTGGGTCTCTGAAAGTCCAACTCGGG 2490
Db 6047 GGGCAGGGCGGTGACGAGCAGCGCTGCGCTCGGTCTGTAAGTCCAACTCGGG 6106
Qy 2491 CACACATGCTGCGCGCGGTGTGGTGGGTGTCATCAAGATGCTGATGGCTGCGGGAG 2550
Db 6107 CACACCGAGCGCGCGCGGTCTCCGCTGTCATCAAGATGCTGACGCGATGCGCCAC 6166
Qy 2551 GGGGTGTTGCGCGGACGCTGTCATGATGAAGCGTTCGCGACGTCGATGCTGCTCGCG 2610
Db 6167 GGAATGCTGCGAAGACGCTGACGTCGACGAGCCCTCGACCATGATGATGCTGCGGT 6226
Qy 2611 GGGCGGTGCGGCTCTGACGAGCGGTGCGTGGCGCGGAGACGCGGAGGCGGTG 2670
Db 6227 GGGCGGTGGAATCTCTCACGAGCGCTCGACTGGCGGAGGAGGAGGAGGCGGCGGTG 6286
Qy 2671 CGGCGGGCGGAGTGTCTGTTGCGGATCGGCGGACGAAATGCGCATGTGATTTGAG 2730
Db 6287 CGCGGGCGCGGTCTCTCTTCGGGATCAGCGGACCAATGCGCATGTGTTGCTCGAA 6346
Qy 2731 GAGGCGCGCGCGGGGGGTGTTGTCGCGGGGTGGGTGTTGGAGGGTCTCCGGGT 2790
Db 6347 GAGGCGCGGTGTTGTCGAGGGTCTTCGTCGTCGAGCGCTCGTTGGCGGGTCCGG 6406
Qy 2791 CTTGCCATTTGCGTGGCTGAGTTCGCTGCGGTTCAGTGGTGTCTCTCGCGCGGTGCT 2850
Db 6407 GTCGCGCGGTGTCAGCGCTTGGGTGTCGCGGAGTTCGCTGCGCGCTCGACGCG 6466
Qy 2851 GAGTGGGTGCGGTGCGGTGCGGTGCGGTCTCTGTCGCGGTGCTGCGGTGCTGAG 2910
Db 6467 CAGATCGAGCGGT-----TGCGCATTCGCTCGCGGATC 6503
Qy 2911 GCTGGTTCGCGGCGGAGCGGTTGCTGATGATGCTGCGAGTCCGCGCGGAGCTT 2970
Db 6504 GTAAGATGACCGCGACCGCGGTGCTGTCAGCG----- 6537
Qy 2971 TCGCTTCGATGTTGGGTGCGGTCTGGCTGTGGCGGGGTGCTCTGAGCATCTGCG 3030
Db 6538 -----GGGCGCTGCTGCTACGTCGCTGACGCGGTGCTCAGTTTCGAGCACCGGCG 6592
Qy 3031 GTCGCTTCGCGCGGACCGGTGAGGAGTGTGTCAGGGGTGGGGGCGCTGCGCGCGGT 3090
Db 6593 GTCGCGCTCGCGCGGGGCGGACGACCTCTGACAGGCGCTGGCGGATCCCGACGCG--- 6649

```
QY 14064 CCTCGCGAAATACACCGCGCCCACTCGCCGCGCATCTCCTACCCCTCACCAGCGCCACCAC 14123
Db      |||||
QY 17673 GCAGGCGAGATCGCGCGCGTAGCTCGCCGGTCCCTCACCCTCGAGAGCGCCCGCG 17732
Db      |||||
QY 14124 CTTATACCCAAACCGCGCACCTCATGAAACATATGCCCCCGGAC---CATGACCAC 14180
Db      |||||
QY 17733 CCGTGTACCTTGCAGCAGCAAGTCCATGCGCCGCCACCTCGCGGCAAGGCGGATGAT 17792
Db      |||||
QY 14181 CCTCAGACACCCCGCCCAACATACACCCACCACTCACCAGCCACGAAAGACCTCGC 14240
Db      |||||
QY 17793 CTCCCTCGCCCTCAGCGGAGAGCACCAGCGAGCGCATCGAGAACCTCCACGGACTGTC 17852
Db      |||||
QY 14241 CATCGCGGCATCAACACCCCGCCCACTCCCTCGTTCATCAGCGGACACCCCGCCACACCGTCCA 14300
Db      |||||
QY 17853 GATCGCGCGCTCAACCGGCCCCACCGGCACCGTGGTTTCGGGGGACCCACCCAGATCCA 17912
Db      |||||
QY 14301 ACATATACACACCTCTGTCACACACAGGCGATCAAAACCAAAACCTTCCCGCCACCAACA 14360
Db      |||||
QY 17913 AGAGCTCGTCAAGGCGTGTGAGGCGGACCGGGTCCGCGCACGGATCATCCCGTTCGACTA 17972
Db      |||||
QY 14361 CGCCTTCACATCCCGCCACCAACCCCATCTCAACCAACTCCACAGCACACCAAAAC 14420
Db      |||||
QY 17973 CCGCTCCACAGCGCCACGTCGAGACCATCGAGAGCGACTCGCGGAGGTCTCGCCGG 18032
Db      |||||
QY 14421 CTTACCTACACCCACACCCCGCTCATACCGGCAACACCCCGGACCAACT 14480
Db      |||||
QY 18033 GCTCAGCGCGGACACCTGAGGTGCGTCTTCTCGACACTCGAAGCGCTGGATCAC 18092
Db      |||||
QY 14481 CTTACCGCCCACTACTGAGCCCAACAGCCCGGCAACCGTGCATAGCGCCACCAACAC 14540
Db      |||||
QY 18093 CGAGCGGTGCTCGAGCGCACCTACTGTGTAACCGCAACCTCCGCCACCGGTTCGCTCGC 18152
Db      |||||
QY 14541 CCAAAACCTCCACCAACAGCGGTACACCTACATCGAACTCGGACCGGACCAACACCT 14600
Db      |||||
QY 18153 CCGCGCGTTCGAGACCTCGCCACGACGAAAGGCTTACCCACTTCATCGAGGTGAGCGC 18212
Db      |||||
QY 14601 CACCACTCTCAGCCAGCAACCTCCCGCAACCCCGCCACCGACCCCTCACCCTCACCCA 14660
Db      |||||
QY 18213 CCAGCGGTCTCACCATGACCTCCCGAGAGCGGTACCGGCTCGGCACCT---CG 18269
Db      |||||
QY 14661 CCGCCAGCACACCCCGCAACCGCTCTCTACCAACCTTCGGCAAAACCAACCAACCTG 14720
Db      |||||
QY 18270 CCGCGAAGAGGAGCGCAGGAGCGTCTGGTACCTCACTCGCCGAAAGCTTGACCAACGG 18329
Db      |||||
QY 14721 GCACCCCACTACACCCACCAACCAACCAACCAACCCACACCCACACCTCGACCT 14780
Db      |||||
QY 18330 CTTACCATGACTGGGCGCCGCTCC---TCCCGCCGAAACCGGCCACACCCCGAGCT 18386
Db      |||||
QY 14781 CCGCACCTACCCCTTCCAAACACCACTACTGGGTCCA 14819
Db      |||||
QY 18387 CCGCACCTACGCTTCCAGGCGGTCACTACTGGGTCCA 18425
Db      |||||
```

RESULT 3

```
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438U1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
```

```
i ORGANISM: Streptomyces venezuelae
US-09-836-821-5
Query Match 9.6%; Score 2951.8; DB 9; Length 36778;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8436; Conservative 0; Mismatches 5722; Indels 1361; Gaps 69;

QY 8 GGATGAGCGGCGGGAAGAACCCCGCTCGGCGAGGGAGGTCTCGAGTGGCCGACG 67
Db      |||||
QY 3561 GCGCGGAGCATCTGCGCGCAGCGTGGCGCGCTGAGGGTCTGGTTCGGGGCGTGGCTT 3620
Db      |||||
QY 68 AGGCGGACGCGCGTCTGCTTTTTCGCGGCAAGGCCCCGAATGCGCGGCAATGG 127
Db      |||||
QY 3621 CCGGTGTCGGCGAGTGCGCGTTCGTTCCCGGCGAGGCGACGAGTGGGCGGCATGG 3680
Db      |||||
QY 128 GAAAGGGAATCTTCGACGCTTCGACGCTTCGCGGAGAGCGTCCGCGCTTCGCAAGCG 187
Db      |||||
QY 3681 GTGCGGAATGCTGGAATCTTTCGCGGTGTTTCGCGCGGCGCATGGCGAAATGCGAGGCG 3740
Db      |||||
QY 188 GTTTCGCGCCCTACGTTCGACTGGTTCGCTGAGCAGAGTGTTCGCGGACTTCGCGGACGCTC 247
Db      |||||
QY 3741 CACTCTCCCGTACGTCGACTGGTTCGCTGAGGCGCGTCTGACGCGAGGCCCCCGGTGGCG 3800
Db      |||||
QY 248 CCGGCTGGACCGGGTGGACGCTCTCCAGCGCAACCTGTTTCGCGCTCATGATCTCCCTGG 307
Db      |||||
QY 3801 CCACGCTGGAGCGGTGATGTCGTGACGCTGTGAGCTTCGCGGTTCATGCTCTGCTGG 3860
Db      |||||
QY 308 CCGCGCTTCGCGCTTCGCAAGGGTTCGAGCGTTCGCGGCTCTGGGACACAGCTGGGCG 367
Db      |||||
QY 3861 CTCGCTGTGGCAGCACCAACCGGGTTCGCGCGGCGCATGGCGCACTTCGAGGCGCG 3920
Db      |||||
QY 368 AGATCGCGGACGACCACTCTCGCGGAGCGCTCTCCCTGGCGGAGCGCGACAGCGTGGTGA 427
Db      |||||
QY 3921 AGATCGCGCGCGTACGTCGCGCTGAGCTTCGCGCTGAGCTGAGAGCGCGCTCTGTCGTGA 3980
Db      |||||
QY 428 CGCTTTGAGACGAGCGACAGACAC---CCTTTCGCGGACCGCGCGCTCTCTCCCTGG 484
Db      |||||
QY 3981 CCTTCGCGCAGCAAGTCCATCGCGCCCACTTCGCGCGGCGGATGCTGTCTCCCTCG 4040
Db      |||||
QY 485 CCGCACCGCGGATGAGTCTCTCGCGCAACGCTTCGCTGAGCGAGCGGACCAACCGGCGC 544
Db      |||||
QY 4041 CGTTCGAGCGAGGACGCGCTCTCGAGCGACTTGGCGGGTTCGACG----- 4085
Db      |||||
QY 545 GGCTCGCGCTCGCAGCGCTCAACGAGCCCCCGGAGCAGTCTTTCGCGTTCGCGCGAGG 604
Db      |||||
QY 4086 GGCTGTCTGCTCGCGCTGTGAACCGGCGCCACCGCCACCGTGTCTCCGTCACCCGCTAC 4145
Db      |||||
QY 605 CCGTCGCGGACCTGCTGTCGCGCACTTCACCGCGCGGAGGTTCGCGCGCATGATCCCGG 664
Db      |||||
QY 4146 AGATCGAAGAGCTTGTCTCGGCGGTGTGAGGCGCATGGGGTCTCGTCCGCGGTCTATTCCCG 4205
Db      |||||
QY 665 TGGAGGTTCCCGCGCACTCCCGCTCTGATGTAGCGCATCGAGGACGCGGTCTCGAGCGGC 724
Db      |||||
QY 4206 TCGACTACGCTCCACAGCGCGGAGTTCGAGATCATCGAGAGGAGCTCTCCCGAGGTTC 4265
Db      |||||
QY 725 TGCTGCCCATCACCCCGCGCTCCCGCATCTCCCGCTTCCACTCTCTCGGTGACCGCGCGCC 784
Db      |||||
QY 4266 TCGCGGCGCTCAGCCCGGAGGCTCGCGCGGTTCGCTTCTCTCGACACTTCGAAGGCGCT 4325
Db      |||||
QY 785 GCCTCGACACCCCGGAGCTAGACCGGGGTACTGCTATCCGCAACATGTCGAGCAGCGTCC 844
Db      |||||
QY 4326 GGATCAGGAGCGCGTGTCTCGACCGGCGCTACTGTTACCGCAACCTCGCGCATCTGTTGG 4385
Db      |||||
QY 845 GGTTCGAGCGCGCGCGCG---CTGCTTCTGCGAGCAGGGGCCCCAAGACGTTTCGTCGAGA 901
Db      |||||
QY 4386 GCTTCGCGCGCGCGCTCGAGACCTTCGCGCAACCGAGGCGGTTCACCCACTTCGTCGAGG 4445
Db      |||||
QY 902 TGAGCCCGCACCCGCTGCTGACCATGGGCTTCACAGGAG----- 939
Db      |||||
QY 4446 TCAGGCGCCACCGCGCTCTCACCATTGGGCGGACCGTTCACCGGTCTGGCGACCC 4505
Db      |||||
QY 940 -----CTGCGCGCGGACCTGGGCGGACACACCGGCA 970
Db      |||||
```

Db 15570 CCAGCAGGACTGATCCCTGCGCGACAGGTGCGCTCCCGCGCCCGGACCCCGTCCCGGGCAC 15629
Qy 11950 -----CTGGGCCCTGCTAGGACGTGAAT 11972
Db 15630 CTCGACTCGAATCACTTCATGCGCGCTCGGGCGCTCCAGGAACCTCAAGGGGACAGCGT 15689
Qy 11973 GCATTTGGCGAATGAAGCGAAGTCTCTGGAATACCTCAAGCGCGTCACTTCGGGACCTGGA 12032
Db 15690 GTCCAGGTGAACGAAGAGAAGTACCTCGACTACCTCGGTGCGTCCAGCGCGGACCTCCA 15749
Qy 12033 CCGCACTCCCGCTCGCTGTACAGGTGCTGAGCGTGAAGCAGGAGCCGATCCGATGTT 12092
Db 15750 CGAGGCGCGTGGCGCGCTCGCGAGCTGGAAGGGAAGGCGGCGAGCCGCTGCGATCGT 15809
Qy 12093 GGGGATGGGTGCTGTTACCCAGCGGGGCGACGTCAACCAACGCGACTGTGGCATCTCGT 12152
Db 15810 CGGCATGGCTCGCGCTGCGCGCGCGTTCGCTCGCCGAGGACCTGTGGCGGTGTT 15869
Qy 12153 CAAGTCCCAAGAGGAAGTATCGGGAGTTCCCGAACCGACCTGGAAGTGAACCTGGAGCA 12212
Db 15870 GGGCGGCGCGAGGACCGCATCTCGAGTTCCCGCAGGACCGCGGTGGGACGTGGAGGG 15929
Qy 12213 GCTCTAGGACCGGACCGCGACCGCTCAGGAACCACTTACACGCGAGCGGAGGTTTCT 12272
Db 15930 CTTGTACGACCCCGAACCCCGAGGCCACGGGCAAGAGTTACGCCCGAGGCGCGATTCCT 15989
Qy 12273 CTATGACGGGGCGACTTCGACCGCGCTTCTTCGAGTTGTTCACCGCGTGAAGCGCTGCG 12332
Db 15990 GTACGAGGGGGGAGTTTCAGCGCGACTTCTTCGGGATCTCGCCGCGAGGCCCTCGC 16049
Qy 12333 AATGGAACCGGACGAGCGCTGCTCGAAACCACTTGGGAAACGTTTCCGAACAGGGCGG 12392
Db 16050 CATGGAACCGGACGAGCTCTCTCTCGAGGCTCTCTGGAGGCTTTCGAGCACGCGCG 16109
Qy 12393 AATCGACCGAGGTTCATCGCGGAAGCGGACCGGGTTTTCGTTGGGATCAATCCGGA 12452
Db 16110 GATCCCGGGGCGCACCGCGCGCGGACCTCGGTGCGCGCTTTCACCGGCGTATGTACCA 16169
Qy 12453 GGACTACACACCGGATACACATCAGCGCTCAAAACGAGTCGAGGGCTACTCTGTCAC 12512
Db 16170 CGACTAGCGCACCGCTCTCACCGAT---GTCCGGAGGGCATCGAGGGCTACTCTGGGAC 16226
Qy 12513 TGGGAGCGGGAAGCAATGCGTCAAGCGGTATCTCTCAAACTTGGGCTCGAAGGCC 12572
Db 16227 CGGCAACTCCGCGAGTGTGCTCGGCGCGCTCGGCTACACGCTTGGCTCGAGGGGCG 16286
Qy 12573 TCGGATCATATGACACCGGTGTTCTCTCGCTCGTCTGCGCTGCACTGCGCTGCA 12632
Db 16287 GGGCGTCAAGGTGACACCGCTGCTGCTCTCGTGTGCGCTGCACTGCGCTGCA 16346
Qy 12633 AGCGTCCGCTCGGTGAAATGCAACATGCGCTCGCAGCGCGCGCTCGCTCATGCGCAC 12692
Db 16347 GGGCTGCGAAGGGGAGGTTCATGCGCTTCGCGCGCGGTGACGCTCATGTCGAC 16406
Qy 12693 TCCCTTCGTTTACCGAGTTTCTTCGCGAGCGGGCTTGGCGCGAGAGCGCGGTGCA 12752
Db 16407 GCGCAGCACCTTCGTCGAGTTCAGCGCTCAGCGCGGCTGGCGCGAGCGCGGTGCA 16466
Qy 12753 GGGCTTTTCGGCGGCGGAGCGGACCGCTGCTGTCGAGGTTGGGATGCTCGTGT 12812
Db 16467 GTCTTCTGTCGACGGGCGGACCGGACCAAGTGTGTCGAGGCGGTGCGGCTCTCTCTGT 16526
Qy 12813 GGAGCGGTCTCCGACCGCGCGCAACGCTCACCGTGTCTGCGCGCTGCTCGCGGCGAG 12872
Db 16527 CGAGCGCTGTCGAGCGCGCTCGAAGGGCCATCGATCTTCGCGCGGTGTTCCGGGGCAC 16586
Qy 12873 CGCGGTCAACAGGACCGCGCAAGCAAGCGCTGACCGACCCCAACGCTGTTTCAAAAT 12932
Db 16587 CGCGGTCAACAGGACCGCGCACGAGCGGCTCTACGCTCCGAACGCGCGGTGCGACGA 16646
Qy 12933 CAAAGTCAATCCGAGGCTTTGGGCAACGCAACCTCTCTCCCTGCGATGTCGATGCGGT 12992
Db 16647 GCGGTCATCCGACGTGCGCTGCGGACGCGCGGCTCAAGACCTCCGACGTCGTCGTCGTC 16706

Qy 12993 GGAGGCCCAACGGGACCAACCTTGGCGACCCGATCGAGGCTCAAGCCCTCGTCA 13052
Db 16707 CGAGGCCCAACGGGACCGGCTACCGGCTCGGCGACCCGATCGAGGCGGAGCGCTCATCG 16766
Qy 13053 AGCTTACGCTTACGACCGGCGCCCAACGCGCGCGCTCTTGGCTCGGAACCTCAAGTCCAA 13112
Db 16767 CACGTACGGGACGCGCGCTGACCGGCAACAGCGCTGCGCTCGGCTCGTGAAGTCCAA 16826
Qy 13113 CATCGGCACTTCATGCGCGCTGCGGCTGCGGCGGCTCATCAAGATGTTGATGCGCT 13172
Db 16827 CATCGGACACACCGGCGCGCGCTGCTCTCGGCGCTGATCAAGATGTTGTCAGGCGAT 16886
Qy 13173 GGGGAATGCTCTGCTCGCGCGGACGTTGATGATGATGAGCGCTCGCGCATGTGGACTG 13232
Db 16887 GGGCCACGGGCTCTGCGGAAGACGCTCCAGTGAAGAACGACGACGACGAGTGAATG 16946
Qy 13233 GTCCGCGGCTGCGGTGCTGCTGACGGAAGCGTGTGCTTGGCGCGGCG---GGGAGGG 13289
Db 16947 GTCCGCGGCGCGGTGCTGAGCTGCTCACAGGCGCATGACTGCGCGGACAGGCGGACG 17006
Qy 13290 GCGGCTACGCGGCGGAGGAGTGTCTATTCATTCGCGCTGAGCGGACCAAGCCGACGTCAT 13349
Db 17007 CGGACTGCGGAGGCGCGGCTCTCTCTTCTGCGCTGACGCGGACGAACGCGACGTCGT 17066
Qy 13350 CCTCGAGGAAGCACCGCGCCCAACATCCCGTCAACACACCGCGGACGCTCCCGG 13409
Db 17067 GCTCGAAGAGGCGCGG-----CGGCGGAG 17092
Qy 13410 AGAATACGCCCGACGAGGATGCGGTAGTGGGATGAGGCTGTGCGGCGAGTCCAGG 13469
Db 17093 AGACCCCTGCTCGAGGCGACCGCGCGCTCGAGCGGT-----CGGTGCGCGCGG 17144
Qy 13470 GGTGCGCGGTGCTGCTGCGGCAAGTTCGAGCGCGCGCTGCGCGCGGCGGCGGCGGCG 13529
Db 17145 CTTGTTGCGGTGCTGCTGCGGCAAGCTTCGCGCGGCTGGAAGCGGCGGCGGATCGGAG 17204
Qy 13530 CTTGACGCCCACTCACCGACAC-----CGGCGCTCGACCTCGCGGCTCGGCTA 13583
Db 17205 CTTGCGCGGCTTGGCTCGAGGCGGCTGACGAGCGCGCGGATCGGCGGCGGCTCGCTG 17264
Qy 13584 CACCTCGCCCAACCGCGCGCGGTTCGACCAACCGCGGCGGCTCATCGCGCGGCGGCGG 13643
Db 17265 GGTACTGCGCGCGCGCGCGGCTTCGAGCACCGCGCGCTGCTGCTCGGCGGCGGAC 17324
Qy 13644 CGACACCTTCTGCAAGCACTCCAGGCACTCGCGGAGGCAACCCGCGCGGCTCAT 13703
Db 17325 GGAAGATTTCGCGGCGGCTGACCGGCTCCGGAAGGACTGATACGCGGACGCGGCTC--- 17381
Qy 13704 CCACGACGCGCCCGCGGCGGACCGGAGCGCGGAGCGGAGGAAAGACCGCATTCAT 13763
Db 17382 -----GGAAGTGGCGCGGCTGCGGCTTCTG 17405
Qy 13764 CTGCTCGGACAGGCGCACCAACCGCGCGGATGCGGCGGCTCTTACACACCGGCGG 13823
Db 17406 GTTCCCGGTTCAGGCGACGCGAGTGGCGCGGATGGGCGGCGGAACTCTCTCGAGGTGCGAA 17465
Qy 13824 GGTCTTGGCGCGGCACTCAACGACATCTGACCGCACTCGACCGCGGCGGCTCGACCGG 13883
Db 17466 GGAAGTTTCGCGCGGCGGCTGCGGCGGAGTGGCGGCGGCTCTCTCGGCTATGTCAGTGGTC 17525
Qy 13884 CTTCTCTCCCTCTCTCACCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13943
Db 17526 GCTGGAGGCGGCTGCTCGGCGGCGGCGG-----GGG 17555
Qy 13944 GGGCGCATGCTCGAGGAGCGGCGGCTACGCGCGGCGGCGGCTTTCGCGCTTCCAGGTGCG 14003
Db 17556 GCGCGCGCATGCTGAGCGGCGGCGGCGGCTGCTCCAGCGGCGGCGGCTTTCGCTGTCATGTTTC 17615
Qy 14004 CTTCCACGCGCTCTCACCGGAGGCTACCATCAACCGCGGCGGCGGCGGCGGCGGCGG 14063
Db 17616 GCT---GGCGAAGGCTCTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 17672

QY 9985 CAGGCGTGGCGTGGCAAGGACGCTGGGTGAGACGCTGGTGGCGCTGTGTCGGGGGTG 10044
Db 13444 ACAGGACCGCGACGACCTTCGTACGGGCGGACCGCGCTCTCGGCAAG----- 13492
QY 10045 GTGGTGTGGTGGCGGATGCGCGGTGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 10104
Db 13493 -----GACGAGCTGAAGGTGCGCGCGCTCGGAGTTCGGCGGGGTGCGAAGT 13539
QY 10105 GGTGTGGGTGAGTGTGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 10164
Db 13540 -----CGGGCTTACCCCCGACCTGGCGCGCTGTCCGACGAGTGGCGCGCGCGCC 13591
QY 10165 CGGTTTGGCGGTTCGCGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 10224
Db 13592 CCGCGCGCGTACCGTCTTGGCGCGCTGCGCGCGGTCCCGCGGACGCGCGCGGAG 13651
QY 10225 GAGCGCGCGGTGGATGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 10284
Db 13652 GGTGTACGGGGACGCTGGCGCGGACGCTGGAGTGTCTCAGGCGCTGGTGGCGGACGAG 13711
QY 10285 GAGCATCCGAGCGGTTCCTCTCGACCTCGACACGACACCGGACCGACCTCGAC 10344
Db 13712 CACTCGCGGGACCGCGCTGCTCTCTGGTACCGCGGTGGGTGGGTGGGTGGGTGGGTGG 13771
QY 10345 ACCGCTGCTGGTGGTGGGCG-----GTGGATGGTGGGCGGTGGGCGGCGGTGGTGG 10399
Db 13772 TCGGCGCGCGATGGCGCGGAGGACCTGTGCGACGCGCGCGCTGGGCTCTGTACGG 13831
QY 10400 CGTGTGGTGAAGCGCGAGTTGGCGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 10459
Db 13832 ACCGCGCAGACGAGAAACCGCGCGCTTTCGCGCTTCTCGACCTGGCGGACGACGCGCTCG 13891
QY 10460 GACTTGAGTCATCCGCTGATGTTCCAGCCAGCGGTCCCGTGACACAGGACCGCGCGT 10519
Db 13892 TGTACCGGACCTGCGCGTGGTGTCTCCGACGCGGCGCTGCGGACGAAACCGGAGCTC 13951
QY 10520 CCGAGTGTGCTGCCAGCGCTCGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 10579
Db 13952 GCGTGCACGACGCGACCATAGGCTGGCGCGCTGCGCTGGCGTCCGTCCGG--CCGAGACG 14009
QY 10580 GTGCGAGGTGTTGCGGTGGTGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 10639
Db 14010 GCACCGCGCACCGCGGTGCGCGCGGAGGACGCTGCTGTGACCGCGGCGACCGCGG 14069
QY 10640 TGTGGGTGGCGGTGGCGCGCATCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 10699
Db 14070 GCTTGGCGGACGCTGGTGGCGCGGACGCTGGTGGCGGAGTGGGCGGTACGCGCTGCTGC 14129
QY 10700 TGTGAGCGCGGTGGTGGGATGCTCCGCGTGGGAGGCTGTCGGCGGCGGAGCTGGCGG 10759
Db 14130 TGTGAGCGCGCGGCGACGAGCGCGCGCGCGCGCGCGAGCTCGTGACAGCTGGAGG 14189
QY 10760 GGTGGGCGGAGGTGGGATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 10819
Db 14190 CCGTGGGAGCGGAGCTCGGTGGCGCGGTGCGACGCTGCGCGGACCGCGGAGCGCTCACCG 14249
QY 10820 GGTGCTGGAGGTGTTCTGCGCGGTGTCGCTGAGCGGTGTCGTCATGCGGTGGTGG 10879
Db 14250 CGGTACTCGACGCCATCCCGCGCGGAAACCGCGTCCCGCGGTGTCGTCACGCGGAGGCG 14309
QY 10880 TGTGAGCATGCGAGTGGCTCTCTCACGCGCGGAGGCTGGGACGCGGTGTTGCGCG 10939
Db 14310 TCTCTCCGACGCGACCTCCGTCCTATGACGAGGAGGACGTTGGAAACAGTACTGCGGC 14369
QY 10940 CCAAGTGGATGCGCTCTTTGCTGGATGAGCT-----GACCGGGGTATGGAGCTGT 10993
Db 14370 CCAAGTTCGACGCGCGGTTCCTCTCGAGGAACTCACCTCGACCGCGCGCATGACCTGG 14429
QY 10994 CGCGCTTGTGCTGCTCTCTCGCGCGGGATCTGCGGTGGCGCGCGGCGGCACT 11053
Db 14430 CAGCGTTCGTGATGTTCTCTCGCGCGCGCGCTCTTGGTGGCGCGGCGGCGGCT 14489
QY 11054 AGCGCGCGCAATGCGCTCTGGACGCGCTGGCGGTACCGCGCGCGCGCGCGGCTGTC 11113

Db 14490 ACGCGCGCGCAACGCGCACCTCGACGCGCTCGCGCGCTCGCGCGCGCGGACCGGACTCC 14549
QY 11114 CGGCGGTGCTCGCTGGCGTGGGGCTGTGGGAGAGGCGGCGGATGACCGGGGACCTGG 11173
Db 14550 CCGCGCTCTCGCTGGCTGGGGCTCTGGGCGGAGACAGCGGATGACCGCGGAGCTGG 14609
QY 11174 CCGCACGACGACCGCGGATCATCGCTTCCGCTTCTGCAATCCCATGTCGACCGCGGACG 11233
Db 14610 GCGAGGCGGACTGGCGCGATGAGCGCGCGGCGCATCGCGGATCAGCGACCGCGAGG 14669
QY 11234 CACTGGCGCTCTTGTGATGCGCGCTGTGATGCGCGCGGTGCTGTGCGCGCGGAC 11293
Db 14670 GCATCGCGCTCTCGACGCGCGCTCGCGGACGACCGCGCTCTGTGCGCGCTGCG 14729
QY 11294 TCGGTCCCGCGCGCGCTGCGG----- 11316
Db 14730 GGTTCGACGCGCGCGGCTGCGGAGCGCGCGCGGAAACGACCGCGCGGAAATCCCGCGCG 14789
QY 11317 -----CCCTGTGTCAGGACCTCTGCGCGCGGACCGCGCGCGCGCGCGCGCGCGCG 11363
Db 14790 TCTTCGGGAGCTGTCGCGCGCGGACCGCTCGGGCGCGCGCGCTCGCGCT 14849
QY 11364 CACACCACTACCGGTGTGCGGACAAAGCG----- 11394
Db 14850 CGACGACGCGGAGCG 14909
QY 11395 -----GCCAGCTGACCGCGCGCTGGCGCGCGGACACACGACGACGACGACGCGCGCG 11452
Db 14910 CGGTACGCTCGCGGACCG 14969
QY 11453 TCGCGCTGCTGCGCTGCGACATCGCGACCGCTCTGGGCGGACACGCGCGCGCGCGCGCG 11512
Db 14970 TCGAGTTCGTGTCGCGGAGGTGCGCGGAGTACTCGGCGACGCGCGCGGTCAACGCGATCG 15029
QY 11513 CCGCGACCGCGCGTTCGCGGACCTCGCGCTTGTGATCTCCTCACCGCGCTCGAATACGGA 11572
Db 15030 ACGCGAAGCGGCTTCTCGACCTCGGCTTGTGATCTCCTGACCGCGCTCGAATCTCGCA 15089
QY 11573 ACCGCTCTCGGACGACCGGACTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11632
Db 15090 ACCGCTCAACTCGCGCGGTGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15149
QY 11633 ACCCGACCACTCACCG 11671
Db 15150 GCGCGCGGCTCTGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15209
QY 11672 -----AGCCAAACCGGACAACTGTCG----- 11696
Db 15210 ACGGAGCGGGAACCGGAAACGGAACGGAACGGAACGGAACGGAACGGAACGGAACG 15269
QY 11697 -----CCCGTGTGGGAGCTCGAATACTCGAATCGCGCTCTCGCGCGCTCGACA 11749
Db 15270 AGACGAGCGCGTGTGCGACAACTGACCGCGCTGGAAGGCGCTTGTGTGACGCGCG 15339
QY 11750 AAACGACGAGCGCG---AGCGAAAGAGTCACTCGCGCTGAGTCACTCATGTTGAGG- 11805
Db 15330 TCTCGACGCGCGCGGAGCGGAGTCTCGGAGCACTGGGCTCTCGCGCTCGATGG 15389
QY 11806 -----TGGAAACGCAACCGGAGCTCGGACGCGGAGCGGAGCGGAGCGGAGCGGAG 11845
Db 15390 TCACGGGCGAGACCGGAGCGGAGCGCGCTCGGAGCGCGGAGCGCGCGCGCGCGCGCG 15449
QY 11846 AGGACG----- 11852
Db 15450 CCGAGGACCGCGCTTGGCGCGCGGAGCGGAGCGCGGCGGAGCGGAGCGGAGCGGAGCGG 15509
QY 11853 -----GAAGTTACATCGGCAACAGAGGCTGAGATTTTCAATTTCAATTTCAATTTCA 11905
Db 15510 GAGTCCCGGACTTATGAAACGCTCGGCGGAGGAACTCTTGGCGCTCTCGACGAGGAC 15569
QY 11906 TCGCGCTGTCTGAAACCGGAGCGCTGCGCACTCGCGCGGTATCCG----- 11949

Db 11273 -----ACCGGTTCTTCTTACCGGCGAAGGCGAGTCAGCGCCCGCGCGCC 11317
QY 7951 GCCACGGGCTTACACACACCCACCGCTTTCGCGCGCGCACTCAACGACATCTGCACC 8010
Db 11318 GCGCGTGAGTGTACGACCGGACCCCGTCTTCGCGCGCGGCTCGACGAGATCTCGCC 11377
QY 8011 CACTCGACCCCACTCGACACACCCCTCTCTCCCTCTCTCACCCAAACGACACGAC 8070
Db 11378 CACTCGAGGTCACTCGAATCGCCCTCTCTCGAGTGAT-----GTTGCGGCGGAG 11431
QY 8071 AACGAGGACGCGCGCACTGTCTCAGACAGACCCGCTAGCCAGCCGCGCTTCTCGCC 8130
Db 11432 GCGAGCGGAGCGCGCTGTCTCAGACGAGACCGGTACACGAGTGCCTGTCTCGCC 11491
QY 8131 TTCCAGGTGCGCTTCCACCGCTCTCTCACGACGCTACCAATACACCCCGCACTACTAC 8190
Db 11492 CTGAGGTGCGCTTCTCGGCTCTGTC---GAGAGTGGGCGATGCGGCGCGCGACTG 11548
QY 8191 GCGGACATCTCTCGCGAAATACCGCGCGCCACTCGCGGCACTCTCAACCTTACC 8250
Db 11549 CTCGCTCACTCGGTGCGGAGATCGCGCGCGCACTCGCGGTGTGTCTCGCTCGCC 11608
QY 8251 GAGGCCACCACTCATACCCAAACGCGCACTCTCATGCAACCAATGCCCCC---GSC 8307
Db 11609 GAGCGCGCGCTGTCTCGCGCGCGCGCGCTCATGAGAGCTGCGCGCGCGTGGC 11668
QY 8308 ACCATGACCACTCTCACACACACCCCGCCACCATCACCCACCATCACCGCCACGAA 8367
Db 11669 GCGATGCTCGCGTCAAGCGCGGAGAGAGATCCGCGTGTGGTGGAGCGGAGAG 11728
QY 8368 AAGGAC-----CTCGCCATCGCGGCATCAACACCCCGCACTCTCTCATCAGC 8418
Db 11729 CGGTACCGGAGCTCTGAGCTGCGCGCGCTCAACCGCGCGGAGCGCGCTCTGTGC 11788
QY 8419 GGCACCCCGCACACGTCACACATCACCACTCTGCGCAACCAAGGATCAAAACC 8478
Db 11789 GCGAGCGGAGCGCGCGCGGAGCGGAGGTATGTGTGCGGCGTCCGCGCGAGGAC 11848
QY 8479 AAAACCTCTCCACCAACACGCTTTCACCTCCCGCCACCAACCCCACTCTCAACCAA 8538
Db 11849 CCGCGCTGCGGTCAGCCAGCGCTTCCACTCTCGGCAATGAGCGGATGTCTGACGGG 11908
QY 8539 CTCACACGACACCAACCTCACTACCAACCCCAACCCCACTCTCACTACCGCC 8598
Db 11909 TTCCGCGCGCTCTGAGACGCTGAGTTCCGCGCGCTCTCTGACCGTGTCTCGAAC 11968
QY 8599 AACA-----CCCCACCGACGACTCTCTACCCCGCACTAGTGACCAACAA 8646
Db 11969 GTCAACCGGCTGCGCGCGCGCGAGACCTGTGCGACCCCGAGTACTGGTCCGCGAC 12028
QY 8647 GCGCGCAACACGCTGAGTACGCCACCAACCAACCAACCTCCACCAACGCGGTACCC 8706
Db 12029 GTCCGCGGACCGTCTCGCTTCTGAGCGCGTCTGCTCTCGCGGACCTCGCGCTGGG 12088
QY 8707 ACCTACATCGAATCGGACCGGACAAACCTTCAACACCCCTCACCAACCAACCTCCCC 8766
Db 12089 ACCTGCTGAGTGTGGCGCGCGAGCGGCTCTCACCGCAATGCGCGCGGCGCGCTCGG 12148
QY 8767 AACCCCGCAACCAACCTCACTCAACCAAC----- 8799
Db 12149 GACACCCCGCGGATTCGCGCTGCGCGCTCCCGCTCTCCCGCGGCTCTCCCGCC 12208
QY 8800 -----CCCCACACACACCCCGCGCTCTCCCGCGGCTCTCCCGCC 12268
Db 12209 GACTCGCGCGCGGCTCGCGCCCGCGCTGCTGTGGCGCTGTGTGGCGCGCAAG 12668
QY 8836 CTCGCGAAACCAACCACTGGACCCCGCACTACACCA-----CCACGACCAACAA 8892
Db 12269 CGGTGGAGACCGAGACCGTCTCGCGAGCGCTCTCGGAGGCGCACCGCCACGCGCA 12328
QY 8893 CCCACACCAAC-----CCACTCGACTCTCCCACTAC 8928
Db 12329 CCCGACTGGACCGCTGTGTCGCGGCTCGCGGCGGACCGCGTGGAGCTTGCACGATAC 12388

QY 8929 CCTTTCAACACACCACTACTGCTCGAAAGCACACAGCCCGGTGCGGCAACGTGTCA 8988
Db 12389 TCTTTCCGCGCGACCGCTACTGCTGAGCGCCCGCGCGGCGACACCGCGGTGGACAC 12448
QY 8989 GCG---CCGGACTCGA CCGCAACCAACACCCCTTACTCGGCGCGCACATTTGAACTGGCG 9045
Db 12449 GCGCGCTCGGTCTCGGACCGCGCGACCAACCGCTGCTCGGCGCGGTGTGACGCTTCG 12508
QY 9046 ACTGACGCTGAGCGCTTCTTTCGAGGCGCTGTCTTTCGAGGTGCATCCGTTGGCTGGCT 9105
Db 12509 GACCGGACGCGCTCTGCTCA CCGCGCGCTCTCTCCCTGCGCACCA CCGGTGGCTCGCG 12568
QY 9106 GACCATGCTCGTTCGCGCGCACGCTGTGTCGCGCGCACCTTCTCTGAACTCGCCCTT 9165
Db 12569 GACCAAGCGCTCTCGGAGCGCTCTGCTCCCGCGCGCGGATGCTGCAACTCGCGCG 12628
QY 9166 CATGCGGACCATAGCTGGCTGCGACCGAGTGTGATGAGTGA CCGTATGCGCGCGCTG 9225
Db 12629 CACGCTGCGGAGTCCCGCGTCTGCTGACGTCGCGGAGCTGACCTCTCTTGAACCGCTG 12688
QY 9226 GTGGTTCTGTGATGCGGCTGTGATGTGCAAGTTGGGTTGCGGCTGCGGATCG--- 9281
Db 12689 GTACTGCCCGGACACGCTGGCTGAGCTGCGCGTGA CCGTGGGCGCGCGCGGAGAG 12748
QY 9282 ---GGAGGCGCGCGTTTGTGAGTGTATGTCGCGCGGTGGGAGTGTGTTGG---TGGG 9336
Db 12749 CCGGTGCGAGTTCGCGCGGGA CCGCGCA CCGCGCTCTCCCTCACTCTCGCGCTCGCC 12808
QY 9337 GTGTGCTCGTTCGCGGTGTGACGCTGTATGCTTCGCGGCTGCTGTTGAGGCTGCT 9396
Db 12809 GACGCGCGCGGTACCGCTGTGCTGCA CCGGACCGGCTGCTGCTGCGCACCGG 12868
QY 9397 GCTGTGTGTGTGCTGATGCTGCGGCGGTGTCGCGCGCGGCTGCGGTGCGG 9456
Db 12869 CCGGAGCTTCGCTGCGCGCGCGCGCTGCGGCCAATGTGCGCGCGCGCGCGCGGAG 12928
QY 9457 GTGATGTGATGTGTCCGTGACCGTTGCTGCGGCTGTGTGTTGTTTGGGCGCGGTG 9516
Db 12929 GTGCGCTGAGCGTCTCTACAGCGGCTCA CCGGAAACGCGCTCGCTCTCGCTCGCTG 12988
QY 9517 TTTTCGCGCTGCTGCGGTGTGCGTGTGATGCGGCGGATTTCTGCTGCTGAGGTGTCTG 9576
Db 12989 TTTCAAGGCGCTGAACCGGTGTGCGGTACAGAGGTGAGGTCTTCGCGCA CACTCGCGCTC 13048
QY 9577 CC-----GGAGGAGCGGTGGGCTGATGCGGCTGTTTTGGGCTGCTGATCCGCGG 9624
Db 13049 CCGCGCAACCAATGCGACCGCGCGCGCGCAACCGCGCGCGGAGTGC CGCGCGCG 13108
QY 9625 TTGCTGATGTGTGCTCAGCGCTGTGCTGCTTTCGCGGTGGGACCGGGTTTGGG 9684
Db 13109 CCGCTTACGCAATCCACCCCGCTGCTGCTGACGCTTCTGCTGCA CCGCATCGCGTCTGCG 13168
QY 9685 GAGGCGCGGCTTCGCGGAGGTGTTTCGCGTTCGCGGTGTGTGCGGCTGCTGCTGCTGCT 9744
Db 13169 GGTCTCTGCGACGAGCGCGAGCTGCTCGCGTCCCTTTCACCTGAGCGGTGTCA CCGCTG 13228
QY 9745 CACCGCGCGGTGTGA CCGGTGTGCGGCTGCTGTGCTGCTGCTGCTGCGGCGGCGCGCGG 9804
Db 13229 CACCGCGCGGTGCGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13276
QY 9805 CGTGAGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9864
Db 13277 ACAGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13336
QY 9865 CGTCTGAGTTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9924
Db 13337 CGGCTCAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13383
QY 9925 CCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9984
Db 13384 CGTCCGCGGCTGATGACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13443

Qy	5875	GCCCGGAAAGGAGAGCCAGCTGTTCAGCGCGCCCACTCTCAACCGCGCG-----AGTTG	5925
Db	9125	GCACGGGACAGCGCCACGTCGCGACAGGGCGGGAGCTCCGCGCAAGGGCGGCAACCCCGCTG	9184
Qy	5926	CACAAAACGCTGGCCCATCAGACGTCGGCGGACCAACAGGGCGGCATTCGCTCGAGCTCGTA	5985
Db	9185	GCCGAGGGGTGGCCGCCCGCGGTCCCGCGAGCGGTACGGAGATCCTCTCGGTCTCGTA	9244
Qy	5986	CGAGACCATGTGCGGCAGTGTCTCGGCACCGCGACCCGAAAGCCATCGCGCCGAGCCAG	6045
Db	9245	CGGCGCAGCGCGCGCTGCTCCGGATCGCTTCGCGGAGGAGCTCGCGCGGACCGC	9304
Qy	6046	TGCTTTCGGTGCACTCGGCTTCGATTCATCA CGGCGCTCGAGTTCGGAACCTGCTGATC	6105
Db	9305	GCCTTCAAGAGCATCGCTTCGACTCGCTCGCGGTGTCGAGTTCGCAACAGGCTGACC	9364
Qy	6106	AAGGCAACAGGACTCGCGCTTCTCTGTCTCGCTGTTCGACCAACCGACCCCTGCCAAA	6165
Db	9365	CGGGCAGCCGGGTCTCAGCTGCCCGGACGCTCTGCTTCGACCAACCGACCGCTGGCC	9424
Qy	6166	CTCGCGGTACACCTGCAGAAACCAACTGCGGGGCA CAGCAGCGGAGTCTGC-----	6215
Db	9425	CTCGTGTGCTGCTCCGACGAGTTCTCTCGGTGACGAGAGACGGCGGACGCCCGGCGG	9484
Qy	6216	-----TCCTTCAGCGGCGAGCGGTTCACCGCGAGGCTTCT	6249
Db	9485	TCCGCGCGCTGCCCGCACTGTGCGTGCCTGCGGTGCGCGCGCGCGGCAACGATGCC	9544
Qy	6250	GTCAACGAGCGGATCGGCATCGTTGGCATGGCTGTGCTTTCCCGGCGGAGTGACCTCG	6309
Db	9545	GAGCAGCATCCGATCCGATCGTCGGGATGAGTCCGCTACCCGCGTGACATCCGCGAGC	9604
Qy	6310	GCGGACGACTTCGCGATCTGATCTCCTCCGAGCAGGACGCGCATCGGCGGATTCGCCACC	6369
Db	9605	CCGGAGGACTGTGGCGGATGCTGTCGAGGGCGGCGAGGAGTACAGCCGTTTCCCGACC	9664
Qy	6370	GACCGGGCTGGACCTGAGCAGCTCTACGACCCCGACCCCGACCA CCGCGGACACTGC	6429
Db	9665	GACCGGGCTGGACCTCGACGGCTGTACGACCGCGCTACGACCGCGACCGGTCGGCAGGGCG	9724
Qy	6430	TACACCGAAACGGCGGATTCCTCTACGACGCGAGGCCACTTCGACGCCGGAATCTTCGGC	6489
Db	9725	TACGTCCGAGGGCGGGTTCTTGCAAGCAGCGCGCGGAGTTTCGACCGCGGAGTTCTTCGGC	9784
Qy	6490	ATCAGCCCCCGAAGCCCTCGCATTGGAACCCCGACGACGACTCTCTCTCGAAACCGCC	6549
Db	9785	GTCTCGCGCGCGAGGGCTGGCCATGGAACCGCAGCAGCGGATGCTCTGACGAGTCC	9844
Qy	6550	TGGGAAACCATCGAAACA CGCGGCATCAACCCGCCACACCTCCAGGCA CCCCCACCGGA	6609
Db	9845	TGGGAGCCCTTCGAGCGGGCGGCATCGAGCGCGCATCGAGCGGCATCGTCGCGGACGACACCGT	9904
Qy	6610	GTCTTACCGGCACCAACGGAAGGACTAGGCATTCGCGGTGCAACAACGCGGGCCAGTCA	6669
Db	9905	GTCTTATCGGCTCTCTCTACCAAGACTACGCGGCGCGGTCCGAAGCCCGCGGTGGC	9964
Qy	6670	ACCGATGGTTTCGCACTACCGGAACCGCGGAGGGTTCATCTCGGTGCTATCTCGTAC	6729
Db	9965	GTGGAGGGTTACTGTGTGACCCGACGACGCCGAGGCTCGCTCGGCGGTATTCGCGTAC	10024
Qy	6730	ACGTTTGGTTTTCAGGGTCTCGGTTGTGCGGTGACACGGCTTGTTCCTCGCTGTGGTG	6789
Db	10025	ACCTTCGTTCTCGAAGGGCCCGGACGACCGTGCACACCGCTGCTGCTCGTCTGCTGACC	10084
Qy	6790	GCTTTGATCTGGCTGTAGGGCTTTCGTTGCGGTGAGTGTCTGATGGCGCTTCGCGGG	6849
Db	10085	GCCCTGCACTGCGGTGCGGGCGCTCGCGAGCGGGAGTGCAATGCGCTCGCTCCCGT	10144
Qy	6850	GGTGTGACGCTGATGCTCTCCGGGTGCTTCGTGGAGTTTTTCGGCGAGCGGGGTCTG	6909
Db	10145	GGCGTGGCGATGATGGCGACCCCGCAATGTTCTGTGGAGTTCAGCCGTTCAGCGGGCGTCT	10204
Qy	6910	GCCGCGACGGGCAATTGCAAGGGCTTCTCGGCGCGCGGACGGGACCGGCTGGGGTGAG	6969

[illegible]

Db 7223 GCGGACGGATTCCGGCGCGGATCAITTCGGTGCATACCGGTCCTCCACACCGCGGAGGTC 7282
Qy 3748 CAGCCCTCGGGAGGAGTTGCTGGAGTCTGCGGGGACATCAGCCCGCAGCGTCCGGC 3807
Db 7283 GAGATCATCAGAGCGAGCTCGCCAGGTCTTCGCGGTCTCAGCCCGCAGCGCCCGCGC 7342
Qy 3808 GTGCGGTTCTTCCAGCGGTGGAGGACCTGGTGGACACCAACACCTTGGACCGCGCC 3867
Db 7343 GTGCGGTTCTTCCAGCGTCCGAGGCACTGGATCACCAGGCGCGTCTTCGACGGCACC 7402
Qy 3868 TACTGTTACCGCAACCTGACACCGCGGTCCGTTTTCAGCGATCGCGTCCAGGCGCTGGCG 3927
Db 7403 TACTGTTACCGCAACCTTCGTCACCGGTTCGCGTTCGCGGCTTCGCGGCTTCGAGACCTTGGC 7461
Qy 3928 GATGACGGACACCGCGGTCTTCGTCGAGTTCAGCGCCCGCCACCCCTCGTCCCGGCATC 3987
Db 7462 -----CGTCGAGGAGGCTTCAGGACTTCGTCGAGGTCAGGCGCCAC 7504
Qy 3988 GAAGACACCAACCGAAGACACCGCGAAGACGTCAACCGGATCGGACGCTTCGCGCGGC 4047
Db 7505 CCGGTCTCAACCATGACCTTCGCGGAGACCGTCAACCGGCTTCGCGACCTTCGTCGGA 7564
Qy 4048 GACACGACACCGCGCTTCCTCAGCGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCG 4107
Db 7565 CAGGAGGCGCAAGAGCGTCTGGTCACTCGCTCGCGGCGGTGGGTCAACGGGCTTCGC 7624
Qy 4108 ACACCCACCACTGGCACCACTACACCCACCAACACCAACCAACCCACCGCACAG 4167
Db 7625 GTGGCATGACTTCGCTTCGCGGCAACCGGCTTCGCGC-----7665
Qy 4168 CACTCGA CTGCGCACTACCCCTTCCAAACACGACGACTACTGGTTCGAGAGCTCACAG 4227
Db 7666 -----CGGTCTGCCCACTACGCTTCAGGCGGAGCGGTACTGGCTCGAGAACACTCC 7720
Qy 4228 CCGGCTGCCGATCCGGTTCGGGTGCGGTGCGGTTCGGGTCCCGGTCGCGGCGGCA 4287
Db 7721 GCCGC-----CCTGGCCACCG 7736
Qy 4288 GGAGCTCGCGGCGGAGCGGAGAGTGGAGTTCGCGGTTCTGGACGCGGTGGCGCCCGCAG 4347
Db 7737 GCGACGACTGGCTACCGGATCGACTG-GAAGGCGCTTCGCGGCTCGAGGGTCCGAG 7795
Qy 4348 GACTCGAAACGCTCGGACCACTCGCGGTGCGCGCTTCGCGGCTTCGAGACAGGTG 4407
Db 7796 CGCACCGGCTGTCGCGCGCTGGCTCGCGCTCACCGCGAGGACCACTCCGCGAGGCG 7855
Qy 4408 GTGCGCGACTCTCGCGCTGGACCGCGCAACCAACGACCAAGCCCGGATCAACACTGG 4467
Db 7856 GCGCGGTGCTCACCGCGTGGTTCGACGCGCGGCGGAGGTTCGAGGTGCTGACGCGCGG 7915
Qy 4468 ACCTACAGGAACCTGGAAACCCCTCACCTCCCGCACCAACCAACCCACCAACACC 4527
Db 7916 GCGGACGACCGGTGAGGCTTCGCGCGCTTCGCGGCTCACCGACTGAGC-----7963
Qy 4528 TGCGTCACTCCGATCCCGAAACCCAGACCCACCCCGGATCACCAACATCTCTCAC 4587
Db 7964 -----7963
Qy 4588 AACCTCCACCAACGCGGATCACCCCTCCACCATCCCGCTCACCTCAACCAACCAACCA 4647
Db 7964 -----7963
Qy 4648 CCCCACCACTCCACCAACCTCCACCAACCGGACAAAGCCCAACCAACCAACCAACCA 4707
Db 7964 -----ACCGGT 7969
Qy 4708 GGAGCATCACGGGCTGCT 4767
Db 7970 GACGGCTTCACCGGCTGCT 8000
Qy 4768 CCCCACACCCACCGGACCTCTCTCACTCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 4827
Db 8001 -----ACGGACTCGTACCGCAGGTTCGCTTCGGTTCAGGCGCTTCGCGACGCG 8047

Qy 4828 CACCCACCAACCCCTCTGGTACGCCACCAACCAACCAACCAACCAACCAACCAACCA 4887
Db 8048 GGAATCAAGCGCGCTGTTGGTCCGTACCCAGGGCGCGTCTCTCGTTCGAGACTCTCGAC 8107
Qy 4888 CCGCTCACACACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4947
Db 8108 ACCCGCGCGACCCCGGCGCATGCTCTGGGCGCTTCGGGCGCTTCGGCTCTGAG 8167
Qy 4948 CACCCACCAACCGCGGATCATCGACTCTCCACCAACCAACCAACCAACCAACCAAC 5007
Db 8168 CACCCGAAACGCTGGCGCGGCTCTGTCGACTCTCCCGCGCAACCGATGCGCGGCTC 8227
Qy 5008 CAGCACTCACCACCAACCTCACCAACCCCAACCAACCAACCAACCAACCAACCAAC 5067
Db 8228 GCCACCTCGTCAACCGCATCTTCGGGCGCAACCGGAGGACCAAGATCGGCATCCGAC 8287
Qy 5068 ACGGGACCAACACCGCGCGCTTACCCCAACCAACCAACCAACCAACCAACCAACCAAC 5127
Db 8288 ACCGACTCACCGCGCGCGCTTCGCGCGCAACCGCTTCAACCGAGCTCGGCGCACCGC 8347
Qy 5128 ACCCGCACCCCGGAGACCTCATCAGCGGCGGACCGGCGGCTTCGCCACCCAC 5187
Db 8348 GACTGGCAGCCCAACCGCATCTCTCTCATCCGCGGCAACCGGAGCTTCGGAGCCA- 8406
Qy 5188 CTCACCAACCTCACCAACCCCAACCCCAACCAACCAACCAACCAACCAACCAACCA 5247
Db 8407 --CGCGGACGTTGATGGCCCAACCGAGGCGGAACCTCTCTCTCTCTCTCTCTCTCT 8464
Qy 5248 GCGCGCGCACCCCGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 5307
Db 8465 GCGGAACAGCCCGCGGACCAACCACTCACCGCGAATCACCGCATCGGCGCGCGC 8524
Qy 5308 CTCACATCACCCTCGGACACGACCAACCAACCAACCAACCAACCAACCAACCAACCA 5367
Db 8525 GTCACATCGCGCTGGAGCTCGCGGACCCCGCAACCGCATCGCACCTCTCTCTCTCT 8584
Qy 5368 ATCCCGCGCAACACCCCTCACCAACCGTATCCACCGAGGATCTCTCGACGACGCG 5427
Db 8585 ATCCCGCGGAGACCGCTCACCGCGTCTCTCACCGCGGCGGCTTCGAGACGCG 8644
Qy 5428 ACCCTCAACCACTCACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5487
Db 8645 ATCGTGGACACCTGACCGCGAGGCTCGCGGCGGCGCAACCGTGGAGGCGCGCGC 8704
Qy 5488 GCGCACTCTCTCAACCACTCACCAACCAACCAACCAACCAACCAACCAACCAACCA 5547
Db 8705 GCTCGGTGCTGACGAGTGAACCGGGAACCTCGACCTCGACGCGTTCGTCTCTCTCG 8764
Qy 5548 TCGCGCGCGCGCTTCGGGCGACCGCGCAACCAACCAACCAACCAACCAACCAACCA 5607
Db 8765 TCGGTTCGAGCATCTCTGGGATCCCGGTGAGGCAACTACCGCGCGCAACCGCTTAC 8824
Qy 5608 CTCGACGCGCTTCGCGCACCAACCGCGCAACCAACCAACCAACCAACCAACCAAC 5667
Db 8825 CTCGAGCGCTTCGCGGCTTCGCGCGGCGCAACCGCGCTTCGCGGCTTCGCGGCT 8884
Qy 5668 GGCACCTGGCAAGGAAACGAGTCTGCTGATTCGGCAAGGCGCGCGCATATCTCGACCG 5727
Db 8885 GGACGTTGGAGCGGTGGCGGATGCGCGCGGTGACGCGGTGGCGGCGGTTCGCGCAAC 8944
Qy 5728 CGCGGTTTCGACCCATGTCACCGGTTGGCGCGGCGC-----5768
Db 8945 CACGCGTTCGCGGATGGAACCGGAACTCGCGCTGGCGGCACTGGAGTTCGCGCTCGG 9004
Qy 5769 -----GTCACGAGCGGATCGCGGACCAACCGGCGGCTATGTCGTCTATCGCGGAC 5820
Db 9005 CGGACGAGACCGGATCACCGTTCGCGGACATCGACTGGGACCGCTTCTACCTCGGAT 9064
Qy 5821 ATCGACTGGAGCAAGATCGAACACCTCTCAGACCAACCGGACCTG-----GTGAGCGG 5874
Db 9065 TCCTCGGTTCGCGCGGAGCCCTCTCTCGAGGAGTTCGCGGAGTTCGCGGCGGATCATCGAC 9124

1534 GCCATCGGCGGTTCCCAACCGACCGCGGATGCCCCACGAAACAGCGTCAACGCCACGAC 1593
1519 GCGATCTCGAGTTCCCGCAGGACCGCGCTGGGAGTGGAGGGCTGTACCAACCGGAT 5218
1594 CCACGACCGCGGACGTTCTATCCGACAGGAGGCGGGTTCCTTCAACGCGGGGAC 1653
5219 CCGGAGCACCCCGGACGTCGTACGTCGCCACGCGGCGGTTTCATCAGAAACGTCGCCGC 5278
1654 TTTCAGCGCGGCTTCTTCGGAATCAGTCCACGTCAGGCACTGGCGATGATCCGACGAC 1713
5279 TTTCAGACCGGCTTCTTCGGAATCTCCGCGCGGAGCCCTTCGCAATGACCCGACGAC 5338
1714 CCGCTGCTCTGAGACGTCCTGGAGGCTTCGAGCGCGGCGGAAATCGATCGCTGTCG 1773
5339 CGGCTCCTCTCGAAACCTCTGGGAGCGCTCGAGACCGCGGATCGACCGACCTCC 5398
1774 GTACGGGGTCCCGTACGGGCGTTCCTCGCGGCGCCCTCTCTTCACTACGCGCGGCT 1833
5399 CTGCGGGGACGCGAGTCCGCGCTTCACTGGGGCGATGACCAACGATACGGGGCGAGC 5458
1834 ATGGACACCGCTCGTCCGAGGCGCGCGGACGTCGAGGCGCACATCTCTCAACCGGTACC 1893
5459 CT-----GCGGACCGCGGGGAAGGCTCGACGCTACCTGTGACCGGCAAC 5506
1894 ACGGCGAGCTCTGTCGGGCGGTATCGCTACAGCTTCGGGCTGGAAGCGCGCGATC 1953
5507 ACGGCGAGCTGATGTCGGGCGCGCTCTGTAACACACTCGGCTTGAGGGCGCGCGCTG 5566
1954 ACGGTGACACGGGGTCTCGGATCGCTCGTGAACGCTGATCTGGCGTCCGCTGCTG 2013
5567 ACGGTGGAACGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5626
2014 CGGTCCGGTGAAGTGCAGCTCGGCTGGCGCGGCGCTCTCGGTCACTGCTCAACCTCGC 2073
5627 CGCAAGGCGAGTGCACATGGGCTCGCGCGGCGGCTGGCGGTATGCGCACCGCGG 5686
2074 ATGTTCAATCAGTTCTCCGGACGCGGCTGTCGTTGGAACGAGTGCAGAGCGTAC 2133
5687 ATGTTCTGATCAGCGGACGCGGCTGGCGCGGCGCTGGCGGCGGCTGGAAGGCTTC 5746
2134 TCGGCTGACGCGACCGGCTGGGCGGAGGCGTGGGATGCTGTTGGTGGAGCG 2193
5747 GCGGCTCGGCGACCGGCTGCTGCTGGTCCGAGGCGTGGGCTCTCTCTCTGTCGAGCG 5806
2194 TTGTCGGATCGGTGCGGCTGGGCTGCTGGCGTGTGACGCGGAGTGGGCT 2253
5807 CTGTCGGAACCGCGCGCAACGACACGAGTCTCTCGCGTCTCGCGGCGGCGCTTG 5866
2254 AACAGGACGGTCTGTAATGGCTGACGCGCGCGAACGCTCGGCTCAGAGCGGGTG 2313
5867 AACAGGACGGCGGAGCAACGCGCTCACGGCTCCGAAACGGGCGCTCGCAGCAGCGCTC 5926
2314 ATCCGCGAGGCTTGGGACCGGCGGTTGTCGTTGGCGGATGTCGATGTTGGAGGG 2373
5927 ATCCGCGGCGCTGCGGACCGCGGCTGACGACCTCGAAGTGTGACGCTGTCGAGGCA 5986
2374 CACGCAACGGGACGAGCTGGGCTGATCCGATCGAGGACAGCGGTTGCTCGCCAGCTAC 2433
5987 CACGCAACGGGACGAGCTGGGCTGATCCGATCGAGGCGGCGGCTGATCGCCACTAC 6046
2434 GGGCAG---CGGCGCGGTGACGCGGCTGTCGCTGGGCTCTGTAAGTCCAAACATCGG 2490
6047 GGCAGGGCGCTGACGACGAAACGCGCTGCGCTCGGCTGCTGTAAGTCCAAACATCGG 6106
2491 CACACCATGGCTCGCGGCTGTCGCTGGGCTGTCGCTGGGCTGTCGCTGGGCTGTCG 2550
6107 CACACCATGGCTCGCGGCTGTCGCTGGGCTGTCGCTGGGCTGTCGCTGGGCTGTCG 6166
2551 GGGGCTGTCGCGGAGCTTGTGATGTAAGCTGTCGCGGAGGCTGTCGCTGGGCTGTCG 2610
6167 GGAATGCTCGCGGAGAGCTGTCAGCTGTCAGAGCGCTCTCGACCAAGATGACTGGT 6226
2611 GGGGCGGTGCGGCTGTCGAGGAGCGGCTGTCGCTGGGCGGAGCGCGGCGGCGGTTG 2670

6227 GCGCGGTGGAATCTCTCACCGAGCGCTCGACTGGCGCGGAGACGAGCGCGGCTG 6286
2671 CGGCGGCGGAGTGTCTGCTCGGATCGCGGACGAATGCGCATGTGATTTTGAG 2730
6287 CGCGGGCGCGGCTCTCTCTCGGATCAGCGCACAAATGCGCATGTGCTGCTGAA 6346
2731 GAGCGCGCGGCGGCGGCTGTGTTGCGCGGGTGGGGTGTGAGAGGCTGCTCCGGT 2790
6347 GAGGCGCGGCTGTGTCGAGGCTGCTTGGTCTGAGCGCTGCTGCTGGGCTGGGCTGG 6406
2791 CTTGCGATTTCTGCTGCTGAGTCTGCTGCGCTCAGTGGCTGTCTGCGCGGCTGCT 2850
6407 GTGCGCGGCTGTGAGCGCTTGGGCTGTGCGGAAAGTCTGCTGCGCGCTCGACG 6466
2851 GAGTCTGCTGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2910
6467 CAGATCGAGCGCT-----TGCAGATTCGCTCGCGGATC 6503
2911 GCTGGGTTGCGGCGCAGCGGAGCGGTTGGTCACTAGTACGTGGCAGTCCGCGCGAGCT 2970
6504 GTACGATGACGCGACGCGGCTGCTGCTGACGCG----- 6537
2971 TCGCTTCCGATGCTGGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3030
6538 -----GGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6592
3031 CTGCTCTGCGGCGGACCGTGGAGCTGTGAGGCTGTGAGGCTGTGGGCGCTGCGCGGCT 3090
6593 GTGCGCTGCGGCGCGGCGGACGACCTGTAAGGCGCTGGCGCTGCTGCTGCTGCTG 6649
3091 GAGCGGATCGCGGCTGACCAACGCGCTGATGCGCGGCTGTGAGCGGCGGCTGCTG 3150
6650 -----CTGATACGCGGACGCTTCCGCTGCTCGGCGAGTGGCG 6688
3151 TTCTGTTTCCGAGCAGGCTGGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3210
6689 TTCTGTTTCCCGGCTGAGGCGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 6748
3211 TCTCGGCTTCCCGGCGGATGAGGCTGCGAGGCTGCTGCGCGGCTGCTGCTGCTGCTG 3270
6749 TCGCGGCTTCCCGGCGGCTGAGGCTGCGAGGCTGCTGCGCGGCTGCTGCTGCTGCTG 6808
3271 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3330
6809 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6868
3331 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3390
6869 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6928
3391 GGTATCGAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3450
6929 GGTATCGAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6988
3451 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3510
6989 GCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7048
3511 GCGCG---TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3567
7049 GCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7108
3568 GAGGAGCTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3627
7109 CTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7162
3628 CGCTCCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3687
7163 ACCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7222
3688 GGCACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3747

Db 16412 CCGGAAACGGAGGCCAGAGCGTCTGTGTCACCTCACTCGCCGAAAGCCTGGACCAACCG 16471
Qy 14721 GCACCCCCACCACTACACCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCTCGACCT 14780
Db 16472 CCTCACCATCGACTGGGCGCGCTCC---TCCCAACCGCAACCGGCAACCAACCCGAGCT 16528
Qy 14781 CCCACCTACCTTCCAAACCAACCAACCACTACTGGTCCA 14819
Db 16529 CCCACCTACCTTCCAGGCGCGTCACTACTGGTCCA 16567

RESULT 2
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 9.6%; Score 2951.8; DB 9; Length 36778;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8436; Conservative 0; Mismatches 5722; Indels 1361; Gaps 69;

Qy 8 GGATGACGCGCGGGGAAGAACCCCGCTCGGCGAGGAGGTCTCGGAGTGGCCGACG 67
Db 3561 GGCGGAGCATCTGGCGGAGCGCTGGCGCGCTTGGGCTTGGTCCGGGGGTGGCTT 3620

Qy 68 AGCGGAGCGCGCGCTGCTCTTCTGTTTTCGCGGAGGCGCGCAATGCGCGGCGATGG 127
Db 3621 CGGCTGCGGCGAGTGGCGTTCGTTTTCGCGGAGGCGCGCAATGCGCGGCGATGG 3680

Qy 128 GAAGGAACTTCTGACGCTTCGACGCTTCGCGGAGAGCGTTCGCGGCGTGGAAAGCG 187
Db 3681 GTGCCAACTGCTGGAATCTTCCGCGGTGTTCCGCGGCGCAATGGCCGAATGCGAGGCG 3740

Qy 188 CGTTCCGCGCTACGTCGACTGCTGCTGAGCAGAGTGTGGGGAATCGCGGAGCGCTC 247
Db 3741 CACTCTCCCGTACGTCGACTGCTGCTGAGCGCGTCTGACGAGGCGCGCTGCTGCTG 3800

Qy 248 CGCGGCTGAGCGCGGTGAGCGTCTGCTCAGCGCAACCTGTTGCGCGCTCATGATCTCC 307
Db 3801 CCACGCTGAGCGGCTGATGCTGCTGAGCGTCTGAGCGTTCGCGCTCATGCTCTCGCT 3860

Qy 308 CGCGCTCTGCGCTCGCAAGGGGTGAGCGTTCGCGGAGAGCGTTCGCGGCGTGGAGAC 367
Db 3861 CTCGCGTGTGGAGCAACCGGCGGTGACCGCGGCGTTCGCGGCGGCGTTCGCGGCGG 3920

Qy 368 AGATCGCGCAGCCACGCTCGGAGGCTGTCTCGCGGCGGCGGCGGCGGCGGCGGCG 427
Db 3921 AGATCGCGCGGCTACGTCGCGGCTGCTGAGCTGAGCGGCGGCGGCGGCGGCGGCG 3980

Qy 428 CGCTTTGAGCCAGGCGACAGACAC---CTTTGCGGCGGCGGCGGCGGCTGCTCGCTCG 484
Db 3981 CCTGCGCAGCAAGTCCATCGCGCGCCACTTCGCGGCGGCGGCGGCGGCGGCGGCGG 4040

Qy 485 CGCGCAGCGCGATGAGTCTTCCCGCGGAACTGCTCGCTGGAGCGGAGCAACCGCGCGC 544
Db 4041 CGCTGAGCGAGGACGCGCTCTTGGAGCGAGCTGCGCGGCTTGGTGGCGGCGGCGG 4085

Qy 545 GGCTCGCCCTCGCAGCGCTCAACGGACCCCGGAGCACAGTCTGTTTCCGCTGCCCCGAGG 604
Db 4086 GGCTGTCCTCGCGCTGTGAGAGGGGCCACCGCCACCGTGTCTCGGTGACCCCGTAC 4145
Qy 605 CGGTGCGGACCTGTGTGGCGGACCTCACCGCGCGCAGGTGGCGACGCGCATGATCCCG 664
Db 4146 AGATCGAAGAGCTTGTCTCGGCGGTGTGAGGCGGATGGGGTCCGTGCGCGGTCAITCCCG 4205
Qy 665 TGGAGTTTCCCGCGCACTCCCTCTGATGTAGCCATCGAGNACGGGTCTGTCAGCGGC 724
Db 4206 TCGACTACGCTCCACAGCGCGAGTTCGAGATCATGAGAGCGAGTTCGCGAGGTCC 4265
Qy 725 TGTGTCCTCATCCCCCAGCGCTCTCGCATCCCTTCCACTCTCTCGGTGACCGCGGCG 784
Db 4266 TCGCGGGCTCAGCGCGGAGGTCTCGCGGTGCGCTTCTCTCGACACTCGAAGGCGCT 4325
Qy 785 GCCTGACACCCCGGAGCTAGACGGCGGTACTGTGTACCGCAACATGTGACGACCGTTC 844
Db 4326 GGATCACGAGCGCTGTCTGACGGCGGTACTGTGTACCGCAACCTGCGCATCTGTGTG 4385
Qy 845 GGTTCGAGCGCGCGCGCG---CTGCTTCTGAGAGGGGCGGCGGCGGCGGCGGCGG 901
Db 4386 GCTTCGCGCGCGCGCTGAGACCTTGGCCACCGAGGGCTTCAACCACTTCTGTCGAG 4445
Qy 902 TGAGCGCGCACCGCGTGTGACATGGGCTCTCAGGAG----- 939
Db 4446 TCAGCGCGCACCGCGTCTTCACTATGGCGCTCTCCCGGAGCGCTCACCAGTCTGGGACCC 4505
Qy 940 -----CTCGCGCGGAGCTGGGCGACACCAACCGGCA 970
Db 4506 TCGCTCGGCAACCGCGGTGAGGCGCGCTAGTCTCGCTCTCGCGAAGCATGGGCCA 4565
Qy 971 CGCGCGACACCGTGTATGAGGCGCGCTGCGCGCGCGGCGGCGGCGGCGGCGGCGGCG 1030
Db 4566 ACAGGCTCGCGGTGAGTGGAGCGCGCTCTCTCGCGGCGGCGGCGGCGGCGGCGGCG 4625
Qy 1031 TGACGCTCTCGC-----CCAACTACGGGCGATGTTGAGAGCTGGCGCA 1075
Db 4626 TCCCGACCTACCGTTCAGACCGGCGGCTGCTGGCGGAGATCGAGGCGCTCGCC 4685
Qy 1076 CCACGCTCTCTCGGCGCGCTGACCGCGCTGCTGCCCGGCGGCGGCGGCGGCGGCGG 1119
Db 4686 CGCGGCGGCGGCGGCGGCTGAGCGCGCTCTCGCGAGGCGGCGGCGGCGGCGGCGGCG 4745
Qy 1120 -----CAGAGTCTGCTCTGAGCTGCTGCGCGCGCGCGCGCGGCGGCGGCGGCG 1174
Db 4746 AGCTCGACCGGAGCGAGCGTTCGCTGATCTTGGACAAGGTTCGCGGCGGCGGCGGCG 4805
Qy 1175 ACGACGGAAACGAGCGCACCGCTCGGATGCGCGCGCGCTCGCGCGGCTTTCGCCCTCG 1234
Db 4806 AGGTCTGGGTACCGGACAGCGGCGGAGATCGAGGTGACCGGAGCTTTCGTCGAGCGCG 4865
Qy 1235 GCTTCGACTCTGATGCGTGTGCAACTGCGCAACCGCTCAGCAAGGCGGCGGCGGCTG 1294
Db 4866 GTTGACCTCTCGAGCGGCGTGGACCTGCGCAACCGGATCAACGCGCTTTCGCGGTAC 4925
Qy 1295 GGTTCGCGTGTGAGCTGATCTTTCGACACACACCGCGCGCGGCTGCGCGCGGCGCTTC 1354
Db 4926 GGATGCGCGCTTCTGATGATCTTTCGACTTCCCGCGCGGCGGCTTTCGCGGAGCGTGC 4985
Qy 1355 GGACCGCGCGCTCGCGCGCTCGACGAGGACACCGCGCGCTTACCGGACTCACCGCGG 1414
Db 4986 TCCTCTGTGACG-----GGAGGCGGCGGCGAACCGCGCGGCTGTCGAGCGCGGT 5038
Qy 1415 GCCACGAGGCG--ACGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1473
Db 5039 CGGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5098
Qy 1474 TTCCCGGCGGAGTCCGCTCCCGAAGGACCTGTGGGAGCTGGCGGCTTCGCGGCGGAGAC 1533
Db 5099 CTGCGCGTGGGCTCGCTTCGCGGAGGACCTGTGGCGGCTGTGGCGGCGGCGGCGG 5158

QY 12513 TGGCAGCGCGCAAGCATTTGGTTCAGGCGGTATCTCTAAACTTCGGGCTCGAAGGCC 12572
DB 14369 CGGCAACTCCGGCAGTGTCCCTTCGGGCGCGTTCGGGTACACGCTTGGCTCGAGGGCC 14428
QY 12573 TCGCATCTATTCGACACCGCGTTCCTCTCGTTCGCTGCGCTTCGCAATTCGGCTGCGA 12632
DB 14429 GGCCTTCACGGTCGACACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14488
QY 12633 AGGCTTCGGTTCGGTGAATGACATGGGCTTCGAGGGGCGCTCTCTCTCTCTCTCTCT 12692
DB 14489 GGCCTTCGCAAGGGCGAGTGCATGGCTTCGCGCGGCGGTGACGGTCTATGTCGAC 14548
QY 12693 TCCCTTCGCTTCACCGAGTTCCTTCGCGAGGGGCGCTTCGCGCGAGCGCGGTCGCA 12752
DB 14549 GCCCAGACCTTCGTCGAGTTCAGCGCTCAGCGGGCTTCGCGCGGACCGCGGTGCA 14608
QY 12753 GGGCTTTTCGGGCGCGCGGACCGGACCGGCTTCGCGAGGGTTCGCGGATGCTGCTGCT 12812
DB 14609 GTCTTCTCTGTCGACCGCGCGAGGACACAGCTGGTCCGAGGGCGTTCGCGCTCTCTCT 14668
QY 12813 GGAGCGGCTTCGAGCGCGCGCGCAACGCTCACGCTGCTTCGCGCGCTCTCTCTCTCT 12872
DB 14669 CGAGCGGCTTCGAGCGCGCTTCGCAAGGGCGCATCGGATCTCTCGCGGTTCGCGGCG 14728
QY 12873 CGCGGTCACACGAGCGCGCAAGCAACCGGCTTCGCGCGACCGCAACCGGTCGTTCA 12932
DB 14729 CGCGGTCACACGAGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTCGCGCGCG 14788
QY 12933 CAAGGTCATCTCGCGAGGCTTCGCGCAACGCAACCTCTCTCTCTCTCTCTCTCTCT 12992
DB 14789 GGGCGTCATCTCGAGCTGCTTCGCGGACCGCGGCTTCGCGCGGCGCGCTTCGCG 14848
QY 12993 GGAGCGGCTTCGAGCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTCGCGCGG 13052
DB 14849 CGAGGCGGCTTCGAGCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTCGCGCGG 14908
QY 13053 AGCTACGCTTCAGGACCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTCGCGCG 13112
DB 14909 CAGGTACGGGCGCGGCGCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTCGCG 14968
QY 13113 CATCGGCGCTTCATCGCGGCTTCGCGGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCT 13172
DB 14969 CATCGGCGCTTCATCGCGGCGCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCT 15028
QY 13173 GCGGATGCTCTCGCGCGGCGCGCGCGCGCGCGCGCGCTTCGCGCGGCGCGCTTC 13232
DB 15029 GCGGCGGCGCTTCGCGGAGCGCGCTTCGCGGAGCGCGCGCGCGCGCTTCGCGG 15088
QY 13233 GTCCGCGGCTTCGAGCTGCTGAGCGAGCGGCTTCGCGGCGCGCGCGCGCGCTTCG 13289
DB 15089 GTCCGCGGCGCGGCTTCGAGCTGCTTCGCGGCGCGCGCGCGCGCGCTTCGCGG 15148
QY 13290 GCGGCTTCGCGGCGGCGAGGCTTCATCTTCGCGGCTTCGCGGCGCGCGCGCGCGCT 13349
DB 15149 CGGACTTCGCGGCGGCGCGGCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 15208
QY 13350 CTTGAGGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGCGGCGCG 13409
DB 15209 GCTCGAAGAGGCG 15234
QY 13410 AGAATCAGCGCGCGCGAGGATTCGCGGTAGTGGCGATGAGGCTTCGCGGCGAGTTC 13469
DB 15235 AGACCTCTCTCTCGAGGCGACCGCGCGCGCGCGCGCGCTTCGCGGCGCGCGCG 15286
QY 13470 GGTGCGCGGCTTCGCTGCTGCGCGCAAGTTCGCGAGCGCGCGCGCGCGCGCGCGCG 13529
DB 15287 CTTGCTGCGCGGCTTCGCTGCTGCGCGAAGACTTCGCGGCGCGCGCGCGCGCGCG 15346
QY 13530 CTTGAGCGCGCTTCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13583
DB 15347 CTTGCGCGGCTTCGCTTCGAGGCGCGTTCGCGAGCGCGCGCGCGCGCGCGCGCG 15406
QY 13584 CACCTCTCG 13643

DB 15407 CGTACTGCG 15466
QY 13644 CGACACTTCTTCGAGCACTTCAGGCACTTCGCGGAGCGGAAACCCACCGCGGTCTAT 13703
DB 15467 GAGCATTTTCGCGAGCGCGCTGACCGCTTCGCGAAGGACTGATAGCGGCGACGCGCTC 15523
QY 13704 CCACAGCAGCG 13763
DB 15524 -----GACGCTGGGCGCGGTTCGT 15547
QY 13764 CTGCTTCGAGCAGGGCACCCAAAGCGCGCGCATGCGCCCAACCGGCTTCACCAACCC 13823
DB 15548 GTTCCCGCTCAGGCGACGAGTGGCGCGGATGCGCGCGGCGGCGGCGGCGGCGGCG 15607
QY 13824 CGTCTTCG 13883
DB 15608 GAGGTTCTCG 15667
QY 13884 CTTCT 13943
DB 15668 GCTGAGGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15697
QY 13944 GCGCGCACTGCTTCAGCAGACCGGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCG 14003
DB 15698 CGCGCGCGCGTTCGAGCGGCTCGAGCTCTCCAGCGCGTTCGCGGTCTCTCTCTCT 15757
QY 14004 CTTCT 14063
DB 15758 GCT---GCGCAAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15814
QY 14064 CTTCT 14123
DB 15815 GCGCGCGGATGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15874
QY 14124 CTTCT 14180
DB 15875 CGTCT 15934
QY 14181 CTTCT 14240
DB 15935 CTTCT 15994
QY 14241 CATCT 14300
DB 15995 GATCT 16054
QY 14301 ACACATCT 14360
DB 16055 AGAGCT 16114
QY 14361 CGCT 14420
DB 16115 CGCT 16174
QY 14421 CTTCT 14480
DB 16175 GCT 16234
QY 14481 CTTCT 14540
DB 16235 CGAGCGGCT 16294
QY 14541 CCAAACT 14600
DB 16295 CCGCGCGCT 16354
QY 14601 CACCACT 14660
DB 16355 CCACT 16411
QY 14661 CCGCGCGCT 14720

D	b	12152	GCACCGCGCAGCCGGCGTCCGCCCGCGAGGGGACGGTCTCTGTGACCGCGCGGACCGGGG	12211	Q	y	11633	ACCCACACACCTCACCCACCACTTCCACACACAACTCC	-----	11671	Q	y					
Q	y	10640	TGCTGGGTGCGCGCGGTCATCTGGCTGTGTGTGGGGTGGGGATCTGCTGT	10699	D	b	13292	GCCCGCGGCACTCGCTCTCCACCTTGGAGCGGAGCTGCCGCGCGGCGCTCGGACCGAG	13351	D	b	11672	-----AGCCACAACCGGACAAACGCTGTGCG	-----	11696	Q	y
D	b	12212	GCCTGGGCGGACTGGTCCCGCGGACGCTGGTGGGAGTGGGGGTACGACGCTGCTGC	12271	Q	y	11750	AAACCGACAGCGCC	---AGCGAAAGAGTCAACCTTGGCTGAAGTCACTCATGTGTAGG-	11805	D	b	13352	ACGAGCGGGAAACCGGAACGGGAACGGAACGGGACGACCGGCTCCCGGAGCACCGCG	13411	D	b
Q	y	10700	TGTTGAGCGCGGTGGTCCGATGCTCCGGGTGGAGGATCTTGGCGGCGGAGCTGGCCG	10759	D	b	11697	-----CCCCTGTGGGGAGCTCCGAAATCTGATCCGCTCTCCCGCTTCGACCA	11749	Q	y	13412	AGACGAGCGCTGTGTGCACAACTGACCCCGCTTGGAGGCGCTTGGTGTGACGGGCC	13471	D	b	
D	b	12272	TGTTGAGCGCGGCGGACGGAACCGCGCGCGGACGAGCTGTCAGAGCTGGAG	12331	D	b	11750	AAACCGACAGCGCC	---AGCGAAAGAGTCAACCTTGGCTGAAGTCACTCATGTGTAGG-	11805	Q	y	13472	TCTCGGACGCCCCCGGAGCGAAGATCTCTGGAGACCTCGGCTCTCGCTCGATGG	13531	D	b
Q	y	10760	CGTTGGGGCGGAGGTGGATTTGTTGCTGTGTATGTGGGGGAGCGCGGAGGTGGTCC	10819	D	b	11806	-----TGGAAACGCAACCCAGCATTCGACAGCCGAAAGCGCTGTATG	11845	Q	y	11906	-----TGGAAACGCAACCCAGCATTCGACAGCCGAAAGCGCTGTATG	11845	Q	y	
D	b	12332	CCCTGGGAGCGGAGTCTCGGTGGCGCGTGGAGCTGCGGACCGGAAAGCCCTCACCG	12391	D	b	13532	TCACGGGCGAGACCGGGAACCGGACCGGTCGGGAGCCCGGACGGGCGCGGTCGGCG	13591	D	b	11846	ACGAGA	-----	11852	Q	y
Q	y	10820	GGCTGTGGAGGGTTCCTGCCGGGTGTCGGCTGACGGGTGTCTGTCATGCGGCTGGTG	10879	D	b	13592	CCGAGGACCGGCTTGGCGGCGCGGGAACGGAGCCGGGCGGGGAGTGAAGACCGCGCGG	13651	D	b	11853	-----GAAAGTTCACATTCGGAACAGAGGCTGAGATTTTCAAAATCATTGACAACGACC	11905	Q	y	
D	b	12392	CCGTACTCGAGCCCATCCCGCGGAAACACCGGCTCACCGGGTCTGCCACAGGAGGG	12451	D	b	13652	GAGTCCCGGACTTCATGNAACGCTCGGCGGAGGAATCTTCGGCTCTCTGACAGGACC	13711	D	b	11906	TCGGCTGTCTGTAACCGGAACGCTGCACTCCGCCCTGATCCG	-----	11949	Q	y
Q	y	10880	TGCTGGAAGATGCGATCGCTCTCTACGCCCGAGCGGCTGGGACCGTGTTCGCGG	10939	D	b	13712	CCAGACGGAATGATCCCTGCGGACGCTCGCTCCGCCCGGACCCGCTCCCGGCGAC	13771	D	b	11950	-----CTGGGCCCTGTAGGACGCTGAAT	11972	Q	y	
D	b	12452	TCCTCTCGAGCGCACCTCCGCTCATGACGAGGAGCTGGAACAGTACTGGGC	12511	D	b	11973	GCAATTCGGAATGAGGAGCTCTTGGATACCTCAAGCGGCTCACTCGGACCTTGA	12032	Q	y	13772	CTCGACTCGAATCACTTCATCGCGGCTTGGCGGCTCCAGGAATCAAGGGGACAGCGT	13831	D	b	
Q	y	10940	CCAGGTGGAATCGGCTCTTTTGTGATGAGCT	-----	10993	Q	y	11973	GCAATTCGGAATGAGGAGCTCTTGGATACCTCAAGCGGCTCACTCGGACCTTGA	12032	Q	y	11973	GCAATTCGGAATGAGGAGCTCTTGGATACCTCAAGCGGCTCACTCGGACCTTGA	12032	Q	y
D	b	12512	CCAGGTGGAATCGGCTCTTTTGTGATGAGCT	-----	10993	D	b	13832	GTCACGCTGGAACGAGGAATACCTCGACTAGCTGCTGCTGCCACCGCGGACCTCCA	13891	D	b	12033	CCGACCTCGGCTGCTGACGAGTGTGAGCGGTGAGCAGGAGCGGATTCGCGATGT	12092	Q	y
Q	y	10994	CGGCTTCTGCTGCTCTCTGCGCGCGGGATCTCTGGGCTCGGCGGAGGCAACT	11053	D	b	13892	CGAGCGCGTGGCGGCTCGGAGCTGGAGGCGGAGGCGGCGAGCGGTCGATCGT	13951	D	b	12093	GGGATGCGGTGCTTACCCAGCGGGGCGAGCTCAACCGGAGCTGTGGGATCTCGT	12152	Q	y	
D	b	12572	CAGGTTCTGTCATGTTCTCTCGCGCGCGGCTTCTGGTGGCGGAGGCGGCT	12631	D	b	13952	CGGATGCGCTGCGGCTGCGCGCGGCTGCGCTCGCGGAGACCTGTGCGCGCTGCT	14011	D	b	12153	CAAGTCCAGACGACGCTATCGGGGAGTTCGCGACCGCTGATGGAACCTTGAAGCA	12212	Q	y	
Q	y	11054	ACGCGCGGCAATCGCGCTTGGAACGCTGCGGTACCGCGCGCGGCGCGGCTGTC	11113	D	b	12153	CAAGTCCAGACGACGCTATCGGGGAGTTCGCGACCGCTGATGGAACCTTGAAGCA	12212	Q	y	14012	GGCGGCGGAGAGACGCGATCTCGGAGTTCGCCAGGACCGCGGCTGGGACGCTGGAGG	14071	D	b	
D	b	12632	ACGCGCGGCAACGCCACCTCGACGCTCTGCGCTGGCGCGCGGCGAGCGGACTCC	12691	D	b	12213	GCTCTAGACCGCGGACCCGACCGCTCAGGAACAGATTAACAACGCGGAGCGGAGGTTCT	12272	Q	y	14072	CCTGTAACAACCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGATTCCT	14131	D	b	
Q	y	11114	CGGGGTGCTGCTGGCTGGGGCTGTGGGAAAGGCGACGCGGATGACCGGCGACTGG	11173	D	b	12273	CTATGACGCGGCGACTTCGACGCGGCTTCTTGAGTGTGTCAACCGCTGAGGCGCTGGC	12332	Q	y	14132	GTAAGGCGGCGGAGTTCGAGCGCGGACTTCTTCGGGATCTCGCGCGCGAGGCGCTCGC	14191	D	b	
D	b	12692	CGGCTCTCTCGCTGGGGCTCTTGGCGCGGAGACGAGCGGATGACCGGCGACTCG	12751	D	b	12333	AATGGAACCGGAGCGGCTGCTGCTGGAACAACCTTGGGAAAGCTTCGAAACGAGCGG	12392	Q	y	14192	CATGGAACCGGAGCAGGCTCTCTCTCTGGAGGCTCTCTGGAGGCTCTCGAGGACGCGCG	14251	D	b	
Q	y	11174	CGGCAACGAGCAGCGGCGATCATCGTTCCGGTCTGATCCATGTCGACCCCGGAG	11233	D	b	12393	AATGGAACCGGAGGCTCTCTCTCTGGAGGCTCTCTGGAGGCTCTCGAGGACGCGCG	12452	Q	y	12393	AATGGAACCGGAGGCTCTCTCTCTGGAGGCTCTCTGGAGGCTCTCGAGGACG				

QY 8539 CTCACGACGACACCAACCTCCTACCTACCAACCCACCAACCCCTCATCATCCGCC 8598
DB 10051 TTCCGCGCGTCTGAGAGCGTGGAGTTCCGGCGCCCTCCCTGACCGTGTCTCGAAC 10110
QY 8599 AACA-----CCCCACCGACAACTCCTCACCCCTCCCACTACTGAGACCCAA 8646
DB 10111 GTACCGCGCTGCGCCCGCGCCGACGACTGTGGACACCCGAGTACTGGTCCGGCAC 10170
QY 8647 GCCCGGAACACCGTACGACACCAACCAACCAACCAACCAACCAACCAACCAACCA 8706
DB 10171 GTCCGGGACCGCTCGCTTCTTCGACCGGCTCCGTGCTTCGCGACCTTCGGCGTGG 10230
QY 8707 ACCTACATGAACTCGGACCCGACCAACACCTCACCACCTCACCACCAACCACTCC 8766
DB 10231 ACCTGCTGAGCTGGGCGCCGACGCGGCTCTCACCGCATGGGGCGGACGGCTCG 10290
QY 8767 AACCCCGCAACCAACCTCCTACCTCACCAC----- 8799
DB 10291 GACACCCCGGAGTTCCGCTGCGGCTCCCGCTCGGCTCTCCCGCGGCTCTCCCGGC 10350
QY 8800 -----CCCCACCAACCCCAACCCCAACCCCAACCTCCTCACAAC 8835
DB 10351 GACTCGCGCGCGCGGCTCCGGCCCGCGCGCTGCTGCTGGCGCTGCTGGCGCCGAG 10410
QY 8836 CTCGCCAAACCAACCACTGACCTGACCCCAACCACTTACACCA-----CAACGAACCA 8892
DB 10411 CGGTGCGAGACCGAGACCGTCCGGAGCGCTCTCGGACGGCGCACCGCCACGCGGA 10470
QY 8893 CCCACACCCACAC-----CACTCGACCTCCCACTAC 8928
DB 10471 CCGGACTGGCAGCGCTGTTGCGCGCTCCGGGCGCACCGCTGGACCTGCCACGTAC 10530
QY 8929 CCTTCCAAACCAACCACTACTGCTCGAAGACACAGCCCGTGCAGGACGTGCA 8988
DB 10531 TCCTTCGGCGGACCGCTACTGCTGGAAGCGCCCGGCGGCGGAGTGGAGAC 10590
QY 8989 GCAG---CCGGACTCGAACCCCAACCCCTACTTCGGCGCCCACTTGGAACTGGCG 9045
DB 10591 GCCGGCTCGTCTCGGACCGCGACCAACCGCTGCTCGGCGCGTGGTCAAGCTTCCG 10650
QY 9046 ACTGACGCTGAGGCTTCTTGAGGCGCTTGTCTTTGAGGTCGATCCGCTGGCTGCT 9105
DB 10651 GACCGGACGCGCTGCTGCTCACGGCGGCTCTCCCTGCGGACCCACCGCTGGCTCG 10710
QY 9106 GACCATCGCTCGCGGACGCTGCTGCTCGGCGCCACCTTCTCGAACTCGCCCTT 9165
DB 10711 GACACCGCTCTGCGGAGCGTCTCTCCCGCGCGCTCTCCCTGCGGACCCACCGCTCG 10770
QY 9166 CATCGCGGACATACGTGGGCTCGACCGAGTGAATGAGCTGACGCTGCAATGCGCGCTG 9225
DB 10771 CACGCTCGGAGTCCGCGGCTCTGCTGAGCTGCGGAGCTGACCTCTTGAACCGCTG 10830
QY 9226 GTGTTCTGTGATGGGGGTGAGTGTGAGGTTGGGGTTGGGCTGGGATGG---- 9281
DB 10831 GTACTGCCGAGCAGCTGGGCTGAGCTGCGGCTGACGCTCGGGGCGCGCGGAG 10890
QY 9282 ---GGAGGCGCGGTTTGTGAGTGTATGCGCGGCTGGAGTCTTGTG---TGGG 9336
DB 10891 CCGGCTGGAGTCCGCGGAGACGGCGACGGCCGCTCTCCCTCCTCCTGCGGCTGCC 10950
QY 9337 GGTGTGCTGCTGGGTGAGTGTGAGCGTGTATGCTCGGGGTGCTGTTGAGGCTGCT 9396
DB 10951 GACGCGCCCGGTACCGCTGCTGCTGACCGACCGCTCTGCTGGCCACCGACCG 11010
QY 9397 GCTGT 9456
DB 11011 CCGGAGCTTCCGCTCGCGCCGACGCTGCGGCCATGTGGCGCGCGGCGCGGAG 11070
QY 9457 GTGATGTGATGT 9516
DB 11071 GTGCCGCTGACAGCTCTTACGAGCGGCTCGACGGGAAACGCTCTGCTTCCGCTG 11130
QY 9517 TTTTCGGGCTGCTGCGGCTGTGCGTGTATGGGGGGGATTTGCTGGCTGAGGTGTGTCTG 9576

DB 11131 TTCCAGGCGCTGAACCGGCTGTGGCGGTACGAGGCTGAGGCTTTCGCCGACATCGCGCTC 11190
QY 9577 CC-----GGAGAGCGGTGGGGTGTATCGGGCTGTGTTTGGGCTGATCGATCCGCG 9624
DB 11191 CCGCGCACCAAGATGCGACCGCCCGGACCGGAAACGGGCGGGAGTGTGGCGGGCG 11250
QY 9625 TTGCTGATGTGTGCTGCTGAGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9684
DB 11251 GCCCTTACGGCATCCACCCCGCTGCTGAGCGCTTCTGCTGACCGCTGCTGCTGCTG 11310
QY 9685 GAGGGGCGGGTTCGGGAGAGGTTTCGGGTGCGGCTGTGTGGGTGTGTGTGTGTGTGT 9744
DB 11311 GGTCTGCTGACGAGCCCGAGCTCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 11370
QY 9745 CACCGGCGGGTGTACCGGCTGTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9804
DB 11371 CACGCGCGGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11418
QY 9805 CGTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9864
DB 11419 ACAGGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11478
QY 9865 CGTCTTGTGAGTTGCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9924
DB 11479 CGGCTCAGGCTGCGGCGGCTCACCGCGATACAGGCGGCGG-----GAGCGG 11525
QY 9925 CCGCGGGTTCGCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9984
DB 11526 CGTGCGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11585
QY 9985 CAGGCGTGGCGCTGCGACAGGACGTGGGTGAGAGCGGTGTGGGCTGTGTGCGGGGTG 10044
DB 11586 ACAGGACCGCACGCGCTTCTGTAAGGCGGACCGCGCTTCTGCGGAG----- 11634
QY 10045 GTGGTGTGCGGTGCGCGATGCGCGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGG 10104
DB 11635 -----GACGAGCTGAGGTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11681
QY 10105 GGTGTGGGTGAGTTGTTGGTGGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGG 10164
DB 11682 -----CGGCTCTTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11733
QY 10165 CGGTTTGGCGGTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10224
DB 11734 CCGGCGCCCGTACCGTCTTCTGCGCGCTGCTGCGCGGCTGCTGCGGCGCGCGGAG 11793
QY 10225 GACGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10284
DB 11794 GGTGTACGGGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11853
QY 10285 GAGCATCCGACCGGTTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10344
DB 11854 CACTCGCGGCGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11913
QY 10345 ACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10399
DB 11914 TCCGGCGCGACGATGGCGCGCGAGGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 11973
QY 10400 CGTGTGTGAGCGGAGTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10459
DB 11974 ACCGCGACAGACGAGAAACCGCGCGCTTCTGCGCTTCTGCGCTTCTGCGCTGCT 12033
QY 10460 GACTTGTGATCATCCGTGATGTTCCAGCCACCGGCTCCGCTGCTGCTGCTGCTGCT 10519
DB 12034 TGTACCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12093
QY 10520 CCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10579
DB 12094 GCGCTGACGACGGACCATCAGGCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 12151
QY 10580 GTCGGAGGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10639

Db 7867 TACGTCCGCGAGGGGGTTCTCTGCAAGACGCGCGCGAGTTTCGACGCGAGTTCTTCGCG 7926
Qy 6490 ATCAGCCCCCGAAGCCCTCGCCATGAGACCCCGCAGCAACGACTCTCTCTCGAAACCGCC 6549
Db 7927 GTCTCGCGCGAGGCGCTGGCCATGGACCCGAGCAGCGGATGCTCTGACGACGTCC 7986
Qy 6550 TGGGAACCATCGAACACGCGGATCAACACCCACACCTTCACGAGACCCCGCAGCA 6609
Db 7987 TGGGAGGCTTCGAGCGCGCGCGCATCGAGCGCGCATCGCTGCGCGCAGCAGCACCGGT 8046
Qy 6610 GTCTTCACCGCACCAAGGACAGGACTACGACTTCGCGTGCAACAGCGGGCGCAGTCA 6669
Db 8047 GTCTTCATCGGCCCTCTCTACAGGACTACGCGCGCGCGCTCCGAAACCGCCCGCGTGGC 8106
Qy 6670 ACCGATGTTTCGCACTGACCGGAACCGCGGAGCGTCTCTCCGCTCGTATCTCTGATC 6729
Db 8107 GTGAGGGTTACTGCTGACCGGAGCAGCGCGAGCGTGGCTCGGCGCGTATCGCGTAC 8166
Qy 6730 ACGTTTGGTTGAGGTTCTGCGGTGTCGCTGGACACGCTTGTCTCTGCTGCTGGTG 6789
Db 8167 ACCTTCGGTCTCGAAGGCGCGCGACGACCGTTCGACACCGCTCTGCTGCTGCTGACC 8226
Qy 6790 GCTTTGATCTGGGCTCTCAGGCGTTGCTGCGGTTGCTGAGTCTCGATGGCGCTTGC CGG 6849
Db 8227 GCCCTGACCTTGGCGGTGCGGCGCTCGCAGCGGCGAGTGCAGATGGCGCTCGCGCGT 8286
Qy 6850 GGTGTACGGTGATGTCTCTCGGCTGCTTCTGAGTCTTCTCGCGCAGCGGGTCTG 6909
Db 8287 GCGTGGCGATGATGGCGACCCCGCACATGTTCTGAGTTCAGCCCTCAGCGGCGCTC 8346
Qy 6910 GCCCGGACGCGGCAATCGAAGGCTTCTCGCGCGCGCGGACGCGCTCGGGTGAG 6969
Db 8347 GCCCGGACGCGCAGCAAGGCTTCTCGCGGACGCGGACGCGGTTTCGCGCGCGGAG 8406
Qy 6970 GGTGTGGGATGCTGTGTGGAGCGCTCTCGACGCCCATTCGACAGGTACCGTGTG 7029
Db 8407 GCGGTGCGCTGTGTCTGTGGAGCGCTCTCGAGCGCGCGGCGCAACGCTCACCGGTTG 8466
Qy 7030 CTGGCCGTGTGCTGTGCGTCAACAGGACGCTGCGAGCAACGCTGACCGCG 7089
Db 8467 CTGCGGTGTGCTGTGCGGTACCGCTCAACAGGACGCGCGCGAGCAAGGCTGACCGCG 8526
Qy 7090 CCACAGCGGCGTCCAGAGCGGTTCATCCGCGAGCGCTTCGCGCAACGCGCGTGTGCG 7149
Db 8527 CCCAACGCAACCTCGCAGCAGCGGTGATCCGCGAGCGCTCGCGAGCGCGCGTGGCA 8586
Qy 7150 GCGGTGATTCGACGCGGTGGAGGCCCAAGGACCGGACCGGACCTTTGGGCGACCGATC 7209
Db 8587 CCCGCGACATCGACGCGCTCGAGACGCAAGGACCGGACCGGACCGTCTCGTGGCGACCCATC 8646
Qy 7210 GAGGCCAGGCGCTCTCGCGACTACGAGACGACCGTCTCGCGGAGGCGCGCTGTGG 7269
Db 8647 GAGGCCAGGCGCTCTCAGGCGACGTAAGGACGAGCGCGCGCGGAAAGCGCGCTCGCC 8706
Qy 7270 CTGGGTGCTCAAGTCAATGTGCTGTCACACAGGCTTCGCGGCGGTGCGCGGGGTG 7329
Db 8707 ATCGGTCTCGTGAAGTCAACATCGGACACACCGCGCGCGCGGTGGCGGCA 8766
Qy 7330 ATCAAGATGATGAGGCTGCGGCTGCGGCTGCTGCTGCGCGGAGGTTGATGTGGATGAG 7389
Db 8767 ATCAAGATGCTCTCGGATGCGCCACCGGACCGCTTCGCGAAGACCGCTTCACGCGGAG 8826
Qy 7390 CCGTCCGCGCATGTGACTGCTCGCGGTGCGGTGCGCTGCTGAGCGGAGCGGTGCC 7449
Db 8827 CCGAGCCGACGCTCGACTGGGCGAACAGCGGCTGGCCCTCTGCTACCGAGCGGATCGAC 8886
Qy 7450 TGGCCGCGGAGGCGGCTTACCGCGGCGAGGAGTGTATCATTTGCGCGTTCAGCGG 7509
Db 8887 TGGCCGCGCGGCGGCTGCG---CGCGCGCGCGCTCTCTCTCTTGGCATCAGCGG 8943
Qy 7510 ACCAAGCCCGCTCATCTCGAAGAGACCGCGCGGACGCTT----- 7554
Db 8944 ACGAAGCGCACGCTGTGCTGGAGCAGGCGCGGATGCTGCTGTGAGGTGCTTGGGCGC 9003

7555 -----CCGGGGGACCAACC 7569
Db 9004 GATGAGTGCTCAGGTGCTGAGACGATAGCATGGCTGGGACGCTGGGACCTCCGAG 9063
Qy 7570 GCCCGGAGGTTGACCGGGGACGAGCATGAGGCTGTGCTCGGCGAGTCTCTGGGTGT-- 7627
Db 9064 GTGCTGAGGCTCTGAGGCTCTGAGGCGCGCGCGCGCGCGCGCGCGCGCTGAGGCGTCC 9123
Qy 7628 -----GGCGGTGGCTGTGGCAAGTTCGACGCGCGCGCGCGCGCGCGCGCG 7674
Db 9124 CTCCCGGGGACCTGCTGCTGGGTGCTGCTCGCAAGACGAGCAGTCTGCTGCGGCGCAG 9183
Qy 7675 GCCCAGGCGCTGACGCGCCACCTCAGCGACCCCGCGCTCGA----- 7718
Db 9184 GCCCGCGCTGACGCGGTGGCTGTGCGACCGCGCGCGCGCGCGCGCGCGCGAG 9243
Qy 7719 -----CTCGCGGATGTCGATACCTCGCCACCGCGCGCGCGCGCGCGCGCG 7770
Db 9244 CCGCGCGCTGCGGAGCGTGGGTACGCTCGCCACGAGCGGTACCGCTTCGCGCAC 9303
Qy 7771 CGCGCACCTCTATCGCGGAGCGGACACGTTCTCTGCAAGCATCTCAGGCACTCGCC 7830
Db 9304 CGCGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9363
Qy 7831 GCAGCGAGCGCCACCGCGCTCATCCAGCAGCGCGCGCGCGCGCGCGCGCGCG 7890
Db 9364 CAGGCGGCACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9414
Qy 7891 GAGCGCGAGGAAGACCGATTCATCTCTCGGACAGGCGACCAACCGCGCGCGCATG 7950
Db 9415 -----ACCGGTCTCTCTTACCGCGCAGGCGAGTCAGCGCGCGCGCGCG 9459
Qy 7951 GCCCAGCGCTTACACACCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8010
Db 9460 GCGCGTGTGCTGTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9519
Qy 8011 CACTCGACCGCGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8070
Db 9520 CACTCGAGGTCACCTCGAATGCGCGCTGCTCGAGTAT-----GTTGCGCGCGAG 9573
Qy 8071 AAGAGGACGCGCGCGCTGCTCGAGCAGACCGCTACCGCGCGCGCGCGCGCGCGCG 8130
Db 9574 GCGAGCGGAGCGCGCGCTGCTCGAGCAGACGCGGTACACGAGTGTGCGCGCTGTCGCC 9633
Qy 8131 TTCCAGTCCGCTTCACCGCGCTCTCGACGAGGCTACCATCACCGCGCGCGCGCGCG 8190
Db 9634 CTGGAGTCTGCGCTCTTTCGCGCTGCTC---GAGAGCTGGGCGATGCGCGCGCGCGCG 9690
Qy 8191 GCGGAGACTCTCTCGCGGAAATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8250
Db 9691 CTGCGTCACTCGGTGCGCGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9750
Qy 8251 GAGCGCACCGCTTCATCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8307
Db 9751 GAGCGCGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9810
Qy 8308 ACCATGACCGCTTCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8367
Db 9811 GCGTGTCTCGCGCTTCAGGCGCGGAGCGAGTCTGCTGCGGTGAGAGCGGAG 9870
Qy 8368 AACGAC-----CTGCGCATCGCGCATCAACACCGCGCGCGCGCGCGCGCGCGCG 8418
Db 9871 CGGTAGCGGAGGCTGTGAGCGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9930
Qy 8419 GCGACCG 8478
Db 9931 GCGAGCGGAGCG 9990
Qy 8479 AAAACCTTCCCG 8538
Db 9991 CGCGCGCTGCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10050

QY 4348 GACTGTGAAACGGTTCGCGACACACACTCGCGTGTGCCCCCTCCGCGCGCTCGAGACGGTG 4407
DB CGACCGGCTGTCCGGCGCTGGCTCGCGTCAAGCGGAGACCACTCCGCGCAGGCC 5997
QY 4408 GTGCGCGCACTTCGCGCTGGCAACGCAACAAACGACGACCGCGCATCAACACTGG 4467
DB GCGCGGTGTCAACCGCGTGTGTCGACGCGCGGCGGAAAGTTCGAGGTGCTGACGCGCGG 6057
QY 4468 ACTACAGAAACCTTGAACACCTTCACTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4527
DB GCGGACGACGCGTGAAGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6105
QY 4528 TGGCTCATCGCCATCCCGAAACCGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4587
DB 6106 ----- 6105
QY 4588 AACCTCACGACGCGCATCAACCGCATCCCGCTCAACCGCGCGCGCGCGCGCGCGCG 4647
DB 6106 ----- 6105
QY 4648 CCCAACACCTCCACACACCTCCACCAACCGCGCAACAGCGCGCGCGCGCGCGCGCGCG 4707
DB 6106 -----ACCGT 6111
QY 4708 GGAGCCATCACCGCGCTGTCT 4767
DB 6112 GACGGCTTACCGCGGTGTCT 6142
QY 4768 CCCACACACCGCGCGCGCT 4827
DB 6143 -----ACCGACTCGTACCGCGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6189
QY 4828 CACCCACACCGCGCGCGCT 4887
DB 6190 GGAATCAAGGCGCGCTGTCT 6249
QY 4888 CCCCTCACACACCGCGCGCGCT 4947
DB 6250 ACCCGCGCGCGCGCGCGCT 6309
QY 4948 CACCCACACCGCGCGCGCT 5007
DB 6310 CACCCGAAACGCTGGCGCGCGCT 6369
QY 5008 CAGCACCTCACCCAAACCTCACCCACCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 5067
DB 6370 GCCCACCCTGTACCGCACTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6429
QY 5068 ACCGCGCACCGCGCGCGCT 5127
DB 6430 ACCGCGCTCAACCGCGCGCGCT 6489
QY 5128 ACCCGCACCGCGCGCGCGCT 5187
DB 6490 GACTGGGACGCGCGCGCGCT 6548
QY 5188 CTACCCACACCTCACCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5247
DB 6549 --CGCGGACGCTGGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6606
QY 5248 GCGCGCGCACCG 5307
DB 6607 GCGGAAACG 6666
QY 5308 CTACCATACCACTCTGGGACCG 5367
DB 6667 GTACCATCTCG 6726
QY 5368 ATCCCG 5427
DB 6727 ATCCCG 6786
QY 5428 ACCCTCACCACTCACCG 5487

DB 6787 ATCTGGGACACGCTGACCG 6846
QY 5488 GCGCACCTCTCTCAACCACTCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 5547
DB 6847 GCTCGGTGTCTGACGAGCTGACCGCGGAGCTGACCGCGGTGCTGCTGCTGCTGCTGCTG 6906
QY 5548 TCCG 5607
DB 6907 TCGGTGTGAGCACTCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6966
QY 5608 CTGCGACCGCGCTCG 5667
DB 6967 CTGCGACCGCGCTCG 7026
QY 5668 GSCACTGCGAAGAAACGCACTCGCTGATTGCGGACAGCGCGCGCGCGCGCGCGCGCGCG 5727
DB 7027 GACCGTGGGAGCGGTGGCG 7086
QY 5728 CGCGGCTTTCGACCG 5768
DB 7087 CAGCGGTGCG 7146
QY 5769 -----GGTCAACG 5820
DB 7147 CGGACGAGACCG 7206
QY 5821 ATGCGCTGAGCAAGATCGAAACACCTCTCAGACGCGCGCGCGCGCGCGCGCGCGCG 5874
DB 7207 TCTTCGCGTGGCG 7266
QY 5875 GCGCGGAAAGGAGCG 5925
DB 7267 GACGCGGACG 7326
QY 5926 CACAAACCGCTCG 5985
DB 7327 GCGGAGCGCTGGCG 7386
QY 5986 CGAGACCATGTGGCG 6045
DB 7387 CGGCG 7446
QY 6046 TCGTTCGCTGCACTCGCGCTTCGATTCACCTCAGCGCGCGCGCGCGCGCGCGCGCGCG 6105
DB 7447 GCTTCAAGGACATCGCGCTTCGACTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7506
QY 6106 AAGGCAACGAGACTCGCGCTTCCTGCTCTCGTGGTTCGACGCGCGCGCGCGCGCGCGCG 6165
DB 7507 CGGCG 7566
QY 6166 CTGCGCGTACCTGCGAGAAACCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6215
DB 7567 CTGCTGCTGCTGCTGCG 7626
QY 6216 -----TCCTTCAGCG 6249
DB 7627 TCCGCGCGCGCTGCG 7686
QY 6250 GTCAACG 6309
DB 7687 GACGACGATCCGATCG 7746
QY 6310 GCGGACGACTTCTGGGATCTGATCTCTCTCGAGCAGGACGCGCGCGCGCGCGCGCGCGCG 6369
DB 7747 CCGGAGGACCTGTGGCG 7806
QY 6370 GACCG 6429
DB 7807 GACCG 7866
QY 6430 TACACCGGAAACG 6489

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 05:09:26 ; Search time 2475 Seconds
(without alignments)
18196.097 Million cell updates/sec

Title: US-09-914-286-1
Perfect score: 30690
Sequence: 1 gtcagaggatgacgacg...acgaagtgagacagctga 30690

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0.

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2956.6	9.6	38506	9	US-09-793-708-19
2	2951.8	9.6	36778	9	US-09-860-846-5
3	2951.8	9.6	36778	9	US-09-836-821-5
4	2951.8	9.6	36778	10	US-09-861-289-5
5	2951.8	9.6	37948	9	US-09-988-384B-5
6	2342	7.6	13842	9	US-09-860-846-30
7	2342	7.6	13842	9	US-09-988-384B-30
8	2342	7.6	13842	9	US-09-836-821-30
9	2342	7.6	13842	10	US-09-861-289-30
10	2304.2	7.5	50937	9	US-09-808-880-1
11	2188.4	7.1	15872	9	US-09-860-846-1
12	2188.4	7.1	15872	9	US-09-988-384B-1
13	2188.4	7.1	15872	9	US-09-836-821-1
14	2188.4	7.1	15872	10	US-09-861-289-1
15	2140.6	7.0	11220	9	US-09-860-846-32
16	2140.6	7.0	11220	9	US-09-988-384B-32
17	2140.6	7.0	11220	9	US-09-836-821-32
18	2140.6	7.0	11220	10	US-09-861-289-32
19	1004.4	3.3	4041	9	US-09-860-846-36

20	1004.4	3.3	4041	9	US-09-988-384B-36	Sequence 36, Appl
21	1004.4	3.3	4041	9	US-09-836-821-36	Sequence 36, Appl
22	1004.4	3.3	4041	10	US-09-861-289-36	Sequence 36, Appl
23	992.4	3.2	4689	9	US-09-860-846-34	Sequence 34, Appl
24	992.4	3.2	4689	9	US-09-988-384B-34	Sequence 34, Appl
25	992.4	3.2	4689	9	US-09-836-821-34	Sequence 34, Appl
26	992.4	3.2	4689	10	US-09-861-289-34	Sequence 34, Appl
27	985	3.2	68750	9	US-10-014-717-1	Sequence 1, Appli
28	637	2.1	15720	9	US-10-025-380-1058	Sequence 1058, Ap
29	637	2.1	15720	10	US-09-922-217-1058	Sequence 1058, Ap
30	637	2.1	15720	10	US-09-833-263-1058	Sequence 1058, Ap
31	481	1.6	5484	9	US-09-712-363-115	Sequence 115, App
32	371.6	1.2	1010	9	US-09-735-056-29	Sequence 29, Appl
33	367	1.2	1973	10	US-09-864-761-3471	Sequence 3471, Ap
34	360.8	1.2	4209	9	US-09-712-363-20	Sequence 20, Appl
35	348	1.1	4851	9	US-09-712-363-116	Sequence 116, App
36	336.8	1.1	1635	10	US-09-864-761-20241	Sequence 20241, A
37	314	1.0	1035	9	US-09-735-056-30	Sequence 30, Appl
C 38	301.8	1.0	12733	9	US-10-032-393-47	Sequence 47, Appl
C 39	301.8	1.0	12739	9	US-10-032-393-8	Sequence 8, Appli
40	272.6	0.9	4830	9	US-09-738-626-3166	Sequence 3166, Ap
C 41	272.6	0.9	3309400	9	US-09-738-626-1	Sequence 1, Appli
42	261	0.9	8730	9	US-10-124-800-1	Sequence 1, Appli
43	253	0.8	615	9	US-10-125-815-3	Sequence 3, Appli
44	253	0.8	925	9	US-09-735-056-1	Sequence 1, Appli
C 45	239.2	0.8	1075	10	US-09-864-761-19241	Sequence 19241, A

ALIGNMENTS

RESULT 1
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 9.6%; Score 2956.6; DB 9; Length 38506;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8439; Conservative
QY 8 GCATGACGGCGGGAGAACCCCGCTCGCGGAGGAGGCTCTCGGAGTGGCCGACG 67
DB 1703 GCCCGACGATCTGCGGCGCGCTGGCGCGCTGAGGGTCTGGTCCGGCGCTGGCTT 1762